

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-069998

(43)Date of publication of application : 07.03.2000

(51)Int.Cl.

C12Q 1/68
C12N 15/09
G01N 27/327
G06T 7/00

(21)Application number : 10-377796

(71)Applicant : AFFYMETRIX INC

(22)Date of filing : 09.12.1998

(72)Inventor : FIEKOWSKI PETER
BARTELL DAN M

(30)Priority

Priority number : 97 69032
97 996737

Priority date : 11.12.1997
23.12.1997

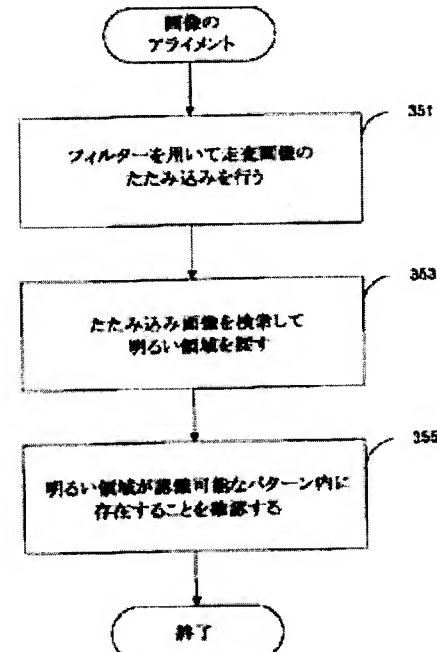
Priority country : US
US

(54) ALIGNMENT SYSTEM FOR SCAN IMAGE AND ALIGNMENT

(57)Abstract:

PROBLEM TO BE SOLVED: To provide an alignment system for folding a scan image including a first pattern by a filter, folding the pattern as a second pattern in the folded image by the filter and carrying out the alignment of a scan image at the position.

SOLUTION: This method for carrying out the alignment of a scan image in a computer system comprises folding a scan image provided with a first pattern composed of a pattern of check, etc., by using a filter, folding the first pattern contained in the scan image as a second pattern composed of a grid pattern, etc., in the folded image by the filter and carrying out the alignment of the scan image according to the position of the second pattern in the folded image so as to carry out the alignment of the scan image by a method for performing the alignment of grid on the scan image of a chip containing a nucleic acid sequence formed in a hybrid.



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-69998

(P2000-69998A)

(43)公開日 平成12年3月7日(2000.3.7)

(51)Int.Cl. ⁷	識別記号	F I	テマコード(参考)
C 12 Q 1/68		C 12 Q 1/68	A
C 12 N 15/09		C 12 N 15/00	A
G 01 N 27/327		G 01 N 27/30	351
G 06 T 7/00		G 06 F 15/62	395

審査請求 未請求 請求項の数1 O.L 外国語出願 (全48頁)

(21)出願番号	特願平10-377796	(71)出願人	598012740 アフィメトリックス・インコーポレーテッド AFFYMETRIX INCORPORATED アメリカ合衆国 カリフォルニア州95051 サンタ・クララ、セントラル・エクスプレスウェイ, 3380
(22)出願日	平成10年12月9日(1998.12.9)	(74)代理人	100096817 弁理士 五十嵐 孝雄 (外2名)
(31)優先権主張番号	60/069,032		
(32)優先日	平成9年12月11日(1997.12.11)		
(33)優先権主張国	米国(US)		
(31)優先権主張番号	08/996,737		
(32)優先日	平成9年12月23日(1997.12.23)		
(33)優先権主張国	米国(US)		

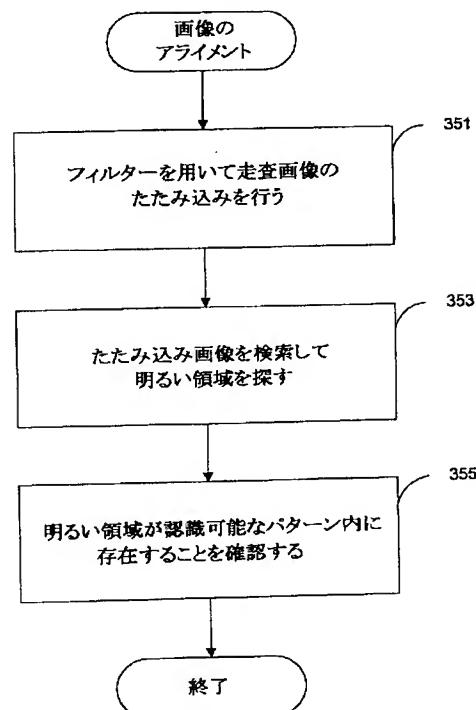
最終頁に続く

(54)【発明の名称】走査画像のアライメントシステム及びアライメント方法

(57)【要約】

【課題】 走査画像のアライメントを行うためのシステム並びに方法を提供する。

【解決手段】 最初に、フィルターを用いて、所定のパターンを含む走査画像のたたみ込みを行うことにより、たたみ込み画像内に認識可能なパターンを形成する。次に、たたみ込み画像における認識可能なパターンの位置に従って、走査画像のアライメントを行う。さらに、フィルターを用いて、走査画像内のパターンに対応しない走査画像の部分を除去することもできる。



【特許請求の範囲】

【請求項1】 コンピュータシステムにおいて、走査画像のアライメントを行う方法であって、
フィルターを用いて、第一のパターンを備えた走査画像のたたみ込みを行う工程であり、前記走査画像に含まれる第一のパターンを、前記フィルターによって、たたみ込み画像内に第二のパターンとしてたたみ込む工程と、
前記たたみ込み画像における前記第二のパターンの位置に従って、前記走査画像のアライメントを行う工程と、
を備える走査画像のアライメント方法。

【請求項2】 前記フィルターを用いて走査画像のたたみ込みを行う工程が、前記第一のパターンの所定の奇数番目の画素と所定の偶数番目の画素との間の相違をたたみ込み画素に設定する工程を備える、ことを特徴とする請求項1記載の方法。

【請求項3】 奇数番目の画素の平均強度が偶数番目の画素の平均強度よりも大きい場合には、前記所定の奇数番目の画素が、奇数番目の画素中で最低の強度を有し、
前記所定の偶数番目の画素が、偶数番目の画素中で最高の強度を有する請求項2記載の方法。

【請求項4】 奇数番目の画素の平均強度が偶数番目の画素の平均強度以下の場合には、前記所定の奇数番目の画素が、奇数番目の画素中で最高の強度を有し、前記所定の偶数番目の画素が、偶数番目の画素中で最低の強度を有する請求項2記載の方法。

【請求項5】 前記第一のパターンが市松模様のパターンである請求項1記載の方法。

【請求項6】 前記第二のパターンがグリッド・パターンである請求項1記載の方法。

【請求項7】 前記走査画像のアライメントを行う工程が、前記走査画像全体にわたってグリッドのアライメントを行う工程を備える請求項1記載の方法。

【請求項8】 前記グリッド内の所定の方向に沿って画素の強度の合計が最小になるように、前記グリッドの位置を調節する工程を、更に備える請求項7記載の方法。

【請求項9】 前記走査画像が、前記第一のパターンの複数のコピーを含む請求項1記載の方法。

【請求項10】 前記走査画像が、それぞれの角の近傍に前記第一のパターンのコピーを有する方形である請求項9記載の方法。

【請求項11】 走査画像のアライメントを行うコンピュータプログラムをコンピュータで読込可能に記録した記録媒体であって、

フィルターを用いて、第一のパターンを備えた走査画像のたたみ込みを行う工程であり、前記走査画像に含まれる第一のパターンを、前記フィルターによって、たたみ込み画像内に第二のパターンとしてたたみ込むコンピュータ・コードと、

前記たたみ込み画像における前記第二のパターンの位置に従って、前記走査画像のアライメントを行うコンピュ

ータ・コードと、
を記録した記録媒体。

【請求項12】 走査画像のアライメントを行う方法であって、

核酸配列を結合したチップで、第一のパターンの核酸配列を含むチップを合成する工程と、

標識された核酸配列と前記チップ上の核酸配列とのハイブリッドを形成する工程と、

前記ハイブリッド形成されたチップを走査して、走査画像を形成する工程と、

フィルターを用いて、前記走査画像のたたみ込みを行う工程であり、前記第一パターンを、前記フィルターによって、たたみ込み画像内に第二のパターンとしてたたみ込む工程と、

前記たたみ込み画像における前記第二のパターンの位置に従って、前記走査画像のアライメントを行う工程と、
を備える走査画像のアライメント方法。

【請求項13】 前記フィルターを用いて走査画像のたたみ込みを行う工程が、前記第一のパターンの所定の奇数番目の画素と所定の偶数番目の画素との間の相違をたたみ込み画素に設定する工程を備える、ことを特徴とする請求項12記載の方法。

【請求項14】 奇数番目の画素の平均強度が偶数番目の画素の平均強度よりも大きい場合には、前記所定の奇数番目の画素が、奇数番目の画素中で最低の強度を有し、
前記所定の偶数番目の画素が、偶数番目の画素中で最高の強度を有する請求項13記載の方法。

【請求項15】 奇数番目の画素の平均強度が偶数番目の画素の平均強度以下の場合には、前記所定の奇数番目の画素が、奇数番目の画素中で最高の強度を有し、前記所定の偶数番目の画素が、偶数番目の画素中で最低の強度を有する請求項13記載の方法。

【請求項16】 前記第一のパターンが市松模様のパターンである請求項12記載の方法。

【請求項17】 前記標識された核酸配列が、前記市松模様のパターンにおける一つおきの方形とハイブリッドを形成する対照核酸配列を含む請求項16記載の方法。

【請求項18】 前記第二のパターンがグリッド・パターンである請求項12記載の方法。

【請求項19】 前記走査画像のアライメントを行う工程が、前記走査画像全体にわたってグリッドのアライメントを行う工程を備える請求項12記載の方法。

【請求項20】 前記グリッド内の所定の方向に沿って画素の強度の合計が最小になるように、前記グリッドの位置を調節する工程を、更に備える請求項19記載の方法。

【請求項21】 前記走査画像が、前記第一のパターンの複数のコピーを含む請求項12記載の方法。

【請求項22】 前記走査画像が、それぞれの角の近傍に前記第一のパターンのコピーを有する方形である請求

項21記載の方法。

【請求項23】走査画像のアライメントを行うコンピュータプログラムをコンピュータで読み可能に記録した記録媒体であって、標識された核酸配列とハイブリッドを形成する核酸配列を結合したチップで、第一のパターンの核酸配列を有するチップの走査画像を入力するコンピュータ・コードと、フィルターを用いて、走査画像のたたみ込みを行う工程であり、前記第一のパターンを、前記フィルターによって、たたみ込み画像内に第二のパターンとしてたたみ込むコンピュータ・コードと、前記たたみ込み画像における前記第二のパターンの位置に従って、前記走査画像のアライメントを行うコンピュータ・コードと、を記録した記録媒体。

【請求項24】第一のパターンでチップに結合された複数のポリマーを備えるチップであって、前記第一のパターンを第二のパターンとしてたたみ込むことにより、走査可能なようにアライメントされるチップ。

【請求項25】前記第一のパターンが市松模様のパターンである請求項24記載のチップ。

【請求項26】前記第二のパターンがグリッド・パターンである請求項24記載のチップ。

【請求項27】前記チップが、それぞれの角の近傍に第一のパターンでチップに結合される前記複数のポリマーを備える方形である請求項24記載のチップ。

【請求項28】前記複数のポリマーが核酸配列である請求項24記載のチップ。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明は、画像処理の分野に関する。さらに詳しくは、本発明は、ハイブリッド形成された核酸配列を含むチップの走査画像上でグリッドのアライメントを行うためのコンピュータシステムに関する。

【0002】

【従来の技術】チップあるいは基質上で試料のアレイを形成し、これを用いるための装置並びにコンピュータシステムが知られている。例えば、本明細書に参考文献として組み込まれるPCT出願WO92/10588及び95/11995に、核酸およびその他の試料の配列決定あるいは配列認識の方法が説明されている。これらの処理を実行するためのアレイを、例えば、本明細書に参考文献として組み込まれる米国特許第5,445,934号及び第5,571,639号に開示される方法に従って形成することができる。

【0003】前記文献に詳述される方法においては、まず、核酸プローブのアレイをチップ上の既知の位置に形成する。次に、標識された核酸をチップに接触させて、

スキャナで、画像ファイル（セルファイルとも呼ばれる）を生成する。この画像ファイルは、標識された核酸がチップに結合している位置を示すものである。画像ファイルと所定の位置におけるプローブの強度とに基づいて、DNAあるいはRNAのヌクレオチド配列あるいはモノマー配列のような情報を抽出することができる。このようなシステムを用いて、例えば、DNAアレイを形成する。DNAアレイを用いることによって、遺伝病、ガン、感染症、HIV（エイズウィルス）等の遺伝特性に関する突然変異を研究し、検出することができる。

【0004】

【発明が解決しようとする課題】VLSIPS（商標）技術を利用することにより、非常に小さなチップ上にオリゴヌクレオチドプローブの非常に大きなアレイを形成することができる。この方法の詳細に関しては、本明細書に参考文献として組み込まれる米国特許第5,143,854号およびPCT特許WO90/15070並びに92/10092に説明されている。DNAプローブアレイ上のオリゴヌクレオチドプローブを用いて、対象となる検体核酸（「標的」核酸）内の相補的な核酸配列を検出する。

【0005】配列の認識を行う場合には、所定の標的核酸配列をチップ上にレイアウトする。例えば、チップが、標的配列に対して完全に相補的なプローブと、標的配列とは一つだけ塩基が異なるプローブとを含むようにしてもよい。新たな配列決定を行う場合には、所定の長さを有する、可能性のある全てのプローブがチップに含まれるようにしててもよい。チップ上で、セルの行と列にプローブを配置する。ここで、各々のセルには、所定のプローブの複数のコピーが含まれる。さらに、プローブをまったく含まない「ブランク」セルをチップ上に設けててもよい。ブランクセルはプローブをまったく含まないため、標識された標的がこの領域内のチップに特異的に結合することがない。従って、ブランクセルを用いて、バックグラウンド強度を測定することができる。

【0006】走査画像ファイルにおいて、一つのセルは、通常、複数の画素で表わされる。走査画像ファイルを視覚的に検査することにより、走査画像ファイル内の個々のセルの同定を行うことができるが、この場合、コンピュータにより実現される画像処理技術を利用して、走査画像ファイルをアライメントすることができれば便利である。

【0007】本発明は、上述した課題を解決するためになされたものであり、走査画像のアライメントを行うための革新的な技術を提供することを目的とする。

【0008】

【課題を解決するための手段およびその作用・効果】上述の課題の少なくとも一部を解決するため、本発明は、以下の構成を採用した。最初に、フィルターを用いて、所定のパターンを含む走査画像のたたみ込みを行うこと

により、たたみ込み画像内に認識可能なパターンを形成する。次に、たたみ込み画像における認識可能なパターンの位置に従って、走査画像のアライメントを行う。さらに、フィルターを用いて、走査画像内のパターンに対応しない走査画像の部分を除去する、即ち、取り除くこともできる。以下に、本発明を実現する実施の形態をいくつか説明する。

【0009】本発明の一つの形態は、走査画像のアライメントを行うためのコンピュータにより実現される方法を提供する。最初に、フィルターを用いて、走査画像のたたみ込みを行う。走査画像には第一のパターンが含まれ、フィルターにより、第一のパターンは、たたみ込み画像内に第二のパターンとしてたたみ込まれる。次に、たたみ込み画像における第二のパターンの位置に従って、走査画像のアライメントを行う。望ましい実施例では、第一のパターンが市松模様のパターンであり、これが、たたみ込み画像内のグリッド・パターンにたたみ込まれる。

【0010】本発明の別の形態は、ハイブリッド形成された核酸配列を有するチップの走査画像のアライメントを行う方法を提供する。最初に、核酸配列（プローブ）を結合したチップを合成する。ここで、チップは、第一のパターンの核酸配列を有する。標識された核酸配列とチップ上の核酸配列とのハイブリッドを形成して、ハイブリッド形成されたチップを走査して、走査画像を生成する。それから、フィルターを用いて走査画像のたたみ込みを行うことにより、第一のパターンをたたみ込んで、たたみ込み画像内に第二のパターンを形成する。次に、たたみ込み画像内における第二のパターンの位置に従って、走査画像のアライメントを行う。望ましい実施例では、第一のパターンが、対照核酸配列によって形成される市松模様のパターンであり、対照核酸配列は、市松模様のパターンにおける一つおきの方形とハイブリッドを形成する。

【0011】本発明の他の特徴や利点に関しては、添付された図面に関連して説明される以下の実施例の詳細な説明により明らかになるであろう。

【0012】

【発明の実施の形態】1. 概要

チップ上に非常に大きなオリゴヌクレオチドプローブのアレイを形成するためのVLSIPS（商標）技術を利用した好適な実施例を参照して、以下に本発明を詳述する。ただし、本発明は、このような方法で形成された画像に限定されるものではなく、他のハイブリッド形成技術や他の技術分野の画像にも好適に適用可能である。即ち、以下の実施例の説明は、何ら本発明を限定するものではなく、例示を目的とするものである。

【0013】図1は、本発明の実施例のソフトウェアを実行するために用いることができるコンピュータシステムの例を示す説明図である。図1に示すコンピュータシ

ステム1は、ディスプレイ3、スクリーン5、キャビネット7、キーボード9、および、マウス11を備える。マウス11には、グラフィカルユーザインターフェースと対話するための一つあるいは複数のボタンが設けられている。キャビネット7には、本発明を実行するコンピュータコードを含むソフトウェアプログラム、本発明に用いられるデータ等を格納し、検索するために用いられるCD-ROMドライブ13、システムメモリ、ハードディスク装置（図2参照）が収容されている。CDS-ROM15がコンピュータ読み取り可能な記憶媒体の一例として示されているが、フロッピーディスク、テープ、フラッシュメモリ、システムメモリ、ハードディスク装置等、他のコンピュータ読み取り可能な記憶媒体を用いてもよい。さらに、例えば、インターネット等のネットワークから取り込むことのできる信号を、コンピュータ読み取り可能な記憶媒体の代わりとして使用することも可能である。

【0014】図2は、本発明の実施例のソフトウェアを実行するために用いられるコンピュータシステム1を示すシステムブロック図である。図1に示すように、コンピュータシステム1は、ディスプレイ3、キーボード9、マウス11を備える。コンピュータシステム1は、さらに、中央処理装置51、システムメモリ53、固定記憶媒体（ハードディスク装置等）55、可動記憶媒体（CD-ROMドライブ等）57、ディスプレイアダプタ59、サウンドカード61、スピーカ63、ネットワークインターフェース65等のサブシステムを備える。ただし、上記以外のサブシステムを備えるコンピュータシステムや、上記のサブシステムのいくつかを省略したコンピュータシステムも、本発明で好適に用いることができる。例えば、コンピュータシステムが、二つ以上の中央処理装置（マルチプロセッサシステム）51やキャッシュメモリを含むようにしてもよい。

【0015】コンピュータシステム1のシステムバスアーキテクチャを矢印67で示す。ただし、これらの矢印は、サブシステムを接続するために用いられる任意の形態の結合スキームを示すものである。例えば、ローカルバスを用いて、中央処理装置をシステムメモリおよびディスプレイアダプタに接続するようにしてもよい。図2に示すコンピュータシステム1は、本発明で好適に用いられるコンピュータシステムの一例を示すものに過ぎず、他の構成のサブシステムを有するコンピュータアーキテクチャも同様に用いることができる。

【0016】本発明は、核酸プローブを含むハイブリッド形成されたチップの走査画像あるいは走査画像ファイルのアライメントを行う方法を提供するものである。本発明の一実施例では、走査画像ファイルには、生物学的アレイから得られる蛍光データが含まれる。あるいは、走査画像ファイルが、放射線強度、光散乱、屈折率、伝導度、エレクトロルミネンス、巨大分子検出データ等

のほかのデータを表わすものでもよい。即ち、本発明は、ハイブリッド形成の蛍光測定値の解析に限定されるものではなく、ハイブリッド形成の他の測定値の解析にも容易に用いることができる。

【0017】以下の実施例では、チップマスクを設計し、チップ上でプローブを合成し、核酸を標識して、ハイブリッド形成された核酸プローブを走査するコンピュータシステムの一部として、本発明を説明する。このようなシステムに関しては、本明細書で参照する米国特許第5,571,639号に詳述されている。ただし、このようなシステム全体とは独立に本発明を用いて、システムで生成されたデータを解析することもできる。

【0018】図3は、RNAあるいはDNA等の生物学的試料のアレイを形成し、解析するためのコンピュータ化されたシステムを示す説明図である。コンピュータシステム100は、RNA及びDNA等の生物学的ポリマーのアレイを設計するために用いられる。コンピュータシステム100は、例えば、適切にプログラミングされたサン社製のワークステーションでもよいし、あるいは、図1および図2に示すように適当なメモリとCPUを備えた、IBM社製のPC互換機等のパーソナルコンピュータやワークステーションでもよい。コンピュータシステム100には、ユーザーが入力した対象遺伝子の特性に関するデータと、所望のアレイの特徴に関するデータが入力される。さらに、GenBank等の外部あるいは内部データベース102から対象となる所定の遺伝子配列に関する情報をコンピュータシステムに入力するようにしてもよい。コンピュータシステム100の出力は、PCT出願WO92/10092に詳述されているようなスイッチ行列の形態のチップ設計コンピュータファイル104並びに他の関連コンピュータファイルの集合である。

【0019】DNA等の分子のアレイを形成する場合に用いられる平版マスクを設計するシステム106にチップ設計ファイルを与える。システム(あるいはプロセス)106には、マスク110を製造するために必要なハードウェア、並びに、マスク上に効率よくマスクパターンを配置するために必要なコンピュータハードウェアおよびソフトウェア108が備えられている。図3に示す他の構成と同様に、この装置を同一の物理的部位に位置するようにしてもよいし、しなくともよいが、図3では、図示の都合上、一緒に示している。システム106は、マスク110あるいはポリマーアレイの製造に際して用いられるガラス上のクロムめっきマスクのような他の合成パターンを形成する。

【0020】マスク110、並びに、コンピュータシステム100から入力されたチップの設計に関する所定の情報が合成システム112で用いられる。合成システム112は、基質あるいはチップ114上でポリマーアレイを形成するために必要なハードウェアおよびソフトウ

エアを備える。例えば、合成システム112が、光源116と、基質あるいはチップ114が載置される化学フローセル118とを備える。マスク110を光源116と基質/チップ114との間に配置した後、チップの所定の領域が露出されるように、適当な回数、光源116と基質/チップ114とを互いに平行移動する。露出された領域への結合、洗浄、その他の操作を目的として、所定の化学薬品を化学フローセル118内に流す。適切にプログラミングされたコンピュータ119の指示に従ってすべての操作を行うことが望ましい。この場合、コンピュータ119は、マスク設計とマスク製造に用いられるコンピュータと同一のものでも異なっていてよい。

【0021】合成システム112によって形成された基質を小さなチップに切断して、標識された標的にさらす。ここで、標的は、基質上の分子の一つあるいは複数に相補的なものでも、あるいは、相補的でないものでもよい。フルオレセイン標識(図3においてアステリスクで示されている)のような標識で標的をマークしてから、走査システム120内に入る。走査システム120は、適切にプログラミングされたデジタルコンピュータ122の命令下で働く。この場合にも、デジタルコンピュータ122が、合成、マスク形成、およびマスク設計の際に用いられたコンピュータと同一のコンピュータでも、あるいは、異なったものでもよい。走査システム120は、標識された標的(*)が基質に結合している位置を検出するために用いられる、共焦点顕微鏡やCCD(電荷結合素子)等の検出装置124を備える。フルオレセインで標識された標的の場合には、基質上の位置の関数として、蛍光強度(フォトン数あるいは電圧等他の関連測定値)を示す画像ファイル124が走査システム120から出力される。標識された標的がポリマーアレイ(例えば、基質上のDNAプローブ)とより強固に結合している部位では計測されるフォトン数が高くなり、また、基質上のポリマーのモノマー配列は位置の関数として周知であるため、標的に対して相補的な基質上のポリマー配列を決定することができる。

【0022】本発明の走査画像アライメント技術を組み込んだ解析システム126に画像ファイル124が入力される。ここでも、解析システムは多様なコンピュータシステムのいずれでもよいが、例えば、WINDOWS・NTワークステーションあるいはそれと同等のものに基づくことが望ましい。解析システムは、画像ファイルを解析して、DNAやRNA等の標的内部における所定の突然変異を同定した結果等の出力128を行う。

【0023】図4は、チップを合成する処理を示す高水準フローチャートである。最初に、ステップ201で、所望のチップ特性をチップ合成システムに入力する。チップ特性には、(配列認識システム等を用いた)対象となる遺伝子配列や標的が含まれる。対象となる配列は、

例えば、ウィルスや微生物、個体を識別するものでもよい。あるいは、対象となる配列が、遺伝死病、ガン、感染症等に関する情報を与えるものでもよい。配列の選択は、テキストファイルの手動入力を介して行ってもよく、あるいは、GenBank等の外部ソースから入力されるものでもよい。新たに標的となる核酸の配列決定を行う場合には、チップには、可能性のある全てのn量体プローブ（ここで、nは核酸プローブの長さを表す）が含まれているため、このステップは不要である。

【0024】新たな配列決定を行う場合には、所定の長さの可能なすべてのプローブを含むセルを備えるようにチップを合成する。例えば、可能な8量体のすべてのDNAプローブを含むように、チップを合成する。このようなチップは、 $65 \times 536 (=4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4)$ 個のセルを備え、各セルが、所定のプローブに対応する。さらに、別の長さを有するすべてのプローブを含む他のプローブを備えるようにチップを合成することもできる。

【0025】ステップ203で、システムは、チップ上でいずれのプローブが望ましいかを判定し、チップ上においてプローブの適切な「レイアウト」を実行する。レイアウトを行うことにより、所望の特性、例えば、遺伝子配列の「読み込み」及び／あるいはエッジ効果の最小化、合成の簡略化等を可能にするチップ上の配置を実現することができる。

【0026】ステップ205で、チップ合成に用いられるマスクの設計を行う。所望のチップ特性およびレイアウトに従ってマスクを設計する。次に、ステップ207

符号	グループ	意味
A	A	アデニン
C	C	シトシン
G	G	グアニン
T	T (U)	チミン (ウラシル)
M	AあるいはC	アミノ
R	AあるいはG	プリン
W	AあるいはT (U)	弱い相互作用 (2つの水素結合)
Y	CあるいはT (U)	ピリミジン
S	CあるいはG	強い相互作用 (3つの水素結合)
K	GあるいはT (U)	ケト
V	A, CあるいはG	T (U) ではない
H	A, CあるいはT (U)	Gではない
D	A, GあるいはT (U)	Cではない
B	C, GあるいはT (U)	Aではない
N	A, C, GあるいはT (U)	コールには強度が不十分
X	A, C, GあるいはT (U)	コールには区別が不十分

大部分の符号は、IUPAC (ユーパック) の標準に従っているが、符号Nは新たに定義され、符号Xが追加されている。

で、システムが、DNAあるいは他のポリマーチップを合成する。ソフトウェアにより、基質とマスクとの相対的な平行移動、フローセルを流れる所望の試薬の流量、フローセルの合成温度、等のパラメータを制御する。

【0027】図5は、DNAプローブ114のアレイに所定の標的DNAが結合した状態を示す説明図である。この簡単な例に示すように、アレイ内では、以下のプローブが形成されている。

3' - AGAACGT
AGACCGT
AGAGCGT
AGATCGT

図示するように、フルオレセインで標識された（あるいはそれ以外の方法でマークされた）標的5' - TCTT GCAをアレイに曝した場合、この標的5' - TCTT GCAは、プローブ3' - AGAACGTにのみ相補的であり、フルオレセインは、主に、3' - AGAACGTが位置するチップの表面上で検出される。チップは、所定のプローブの複数のコピーを含むセルを備えている。従って、画像ファイルには、プローブ（あるいはセル）ごとの蛍光強度が含まれる。所定のプローブに関連する蛍光強度を解析することにより、本明細書に開示される発明の方法を用いて、このようなアレイから配列情報を得ることができる。

【0028】各塩基に割り当てられた以下の符号を用いて、塩基のコール（呼び出し）を行うようにしてもよい。

【0029】2. 走査画像のアライメント
本発明の走査画像のアライメントを説明する前に、実施例の処理全体を簡単に説明する。図6は、チップをハイ

ブリッド形成して、解析することにより、実験結果を得るために処理を示すフローチャートである。ハイブリッド形成ステップ255で、核酸配列（即ちプローブ）が結合したチップ251を、検体の核酸配列（例えば、検体の標識されたフラグメント）と試薬253とに結合させる。ハイブリッド形成ステップにより、ハイブリッドを形成したチップ257が形成される。

【0030】次に、ステップ259で、ハイブリッドを形成したチップを走査する。例えば、ハイブリッドを形成したチップをレーザーで走査して、フルオレセインで標識された検体のフラグメントがチップとハイブリッドを形成している部位を検出する。検体のフラグメントの標識には、様々な手法を適用可能である。また、採用された標識の種類に応じて、走査が行われる。走査ステップにより、チップのデジタル画像261が生成される。

【0031】チップの走査画像が、ハイブリッド形成強度あるいはセル内のプローブに対する検体の親和性に対応する種々の蛍光強度を含むような構成が望ましい。得られた結果の精度を上げるためにには、チップ上の各セルに含まれる画素を同定することが望ましい。画像のアライメントステップ263では、走査画像のアライメントを実行して、各セルに対応する画素の同定を行う。さらに、画像のアライメントステップが、走査画像全体にわたってグリッドをアライメントする処理265を含むようにしてもよい（図7（b）参照）。

【0032】ステップ267で、解析システムが走査画像を解析して、チップ上の各々の対象セルに関する相対的なハイブリッド形成強度を算出する。例えば、セルに関するハイブリッド形成強度、および、セルのプローブと検体の配列との間の相対的なハイブリッド形成親和性を、セルに含まれる画素置の平均として算出することもできる。画素値は、例えば、標識されたハイブリッド形成検体フラグメントから出力されるフォトン数に対応する。

【0033】セル強度を、セル強度ファイル269として記憶するようにしてもよい。セル強度ファイルが、各々のセルに関するセル強度のリストを含むような構成が望ましい。次に、解析ステップ271で、解析システムがセル強度ファイル並びにチップ特性を解析して、結果273を出力する。チップ特性を利用して、チップ上の各セルにおいて合成されたプローブを同定することができる。プローブの配列とセル強度ファイルから読み出されるプローブのハイブリッド形成強度の両方を解析することにより、システムは、突然変異の位置、欠失あるいは挿入、検体の核酸配列等の配列情報を抽出することができる。従って、得られた解析結果は、例えば、配列情報、プローブのハイブリッド形成強度のグラフ、配列間の相違のグラフ等である。これらの詳細に関しては、本明細書に参考文献として組み込まれる米国特許出願第08/327,525号に説明されている。

【0034】走査画像のアライメントを行うために、本発明では、走査画像内に所定のパターンが準備される。この所定のパターンは、認識可能なパターンにたたみ込まれる。望ましい例では、走査画像内に含まれる所定のパターンは市松模様のパターンであって、この市松模様のパターンは対象核酸配列に対して相補的なプローブを含む一つおきのセルを合成することによって形成される。対照拡散配列は、走査画像をアライメントする目的で標識されチップとハイブリッドを形成する周知の配列を用いることもできる。さらに、対照核酸配列に対して相補的なセルの明度を、基準線として、あるいは、他の強度との比較のために用いることもできる。

【0035】一例として、図7（a）に、ハイブリッドを形成したチップに含まれる市松模様のパターンを示す。ハイブリッドを形成したチップの走査画像301には、プローブが合成された活性領域303が含まれる。この活性領域の角に、所定のパターン305として、市松模様のパターンが生じる。通常、所定のパターンは、走査画像の活性領域の各々の角に現れる。ここでは、所定のパターンを市松模様のパターンとしたが、円形、正方形、プラスの符号等、他のパターンを用いることもできる。

【0036】図6を用いて上述したように、チップ上の個々のセルを示すように、走査画像全体にわたってグリッドを配置するようにしてもよい。図7（b）は、図7（a）の走査画像全体にわたってアライメントされたグリッドを示す説明図である。図示するように、ハイブリッドを形成したチップ301の活性領域303全体にわたって、グリッド307が配置される。

【0037】図8は、画像アライメント処理を示すフローチャートである。このフローチャートは、図6のステップ263で実行される処理の詳細を示すものである。まず、ステップ351で、フィルターを用いて、走査画像のたたみ込みを行う。ここで用いられるフィルターは、通常、走査画像をたたみ込み画像にたたみ込むソフトウェアのフィルターである。走査画像のたたみ込みが実施されると、走査画像内の所定のパターンが、認識可能なパターンにたたみ込まれる。たたみ込み画像内の認識可能なパターンの位置を利用して、例えば、画像全体にわたってグリッドを配置する等の方法で、走査画像のアライメントを行うことができる。

【0038】ステップ353で、たたみ込み画像を検索して、明るい領域を探す。走査画像がたたみ込まれると、走査画像に含まれるパターンは、明るい領域を形成する認識可能なパターンにたたみ込まれる。従って、たたみ込み画像内で明るい領域が識別されると、システムは、次に、ステップ355で、明るい領域が予想される認識可能なパターン（例えば、グリッドパターン）内に存在することを確認する。

【0039】パターンによる差を理解できるように、図

9(a)に走査画像403に含まれる市松模様のパターン401を、また、図9(b)にたたみ込み画像453に含まれる認識可能なパターン451を示す。たたみ込み画像は、図9(a)の走査画像から形成される。図示するように、この実施例では、認識可能なパターン451は、フィルターを用いたたみ込みにより市松模様のパターンから形成されたグリッドパターンである。さらに、他の画素強度を除去するようにフィルターが作用することにより、たたみ込み画像には、認識可能なパターンのみが含まれるようになる。走査画像内のパターンの一部ではない画素強度を除去することにより、走査画像をアラインすることが容易になる。

【0040】図10は、走査画像をたたみ込む処理を示すフローチャートである。このフローチャートは、図8のステップ351で実行可能な処理の詳細を示す。最初に、ステップ501で、一つの画素を選択する。この場合、例えば、走査画像の画素を左から右、上から下の順に選択すると仮定してもよい。もちろん、画素を解析する順序を変更することは可能である。

【0041】一つの画素が選択されると、次に、ステップ503で、近隣画素を選択する。近隣画素は、選択された一つの画素に接していないでも、近傍の画素であればよい。例えば、図11に、近隣画素を示す。この近隣画素を解析することにより、たたみ込み画像内にたたみ込み画素が生成される。図11に示す例では、1~9までの番号をつけられた9つの画素がある。例えば、画素1がステップ501で検索された画素であり、ステップ503で検索された近隣画素が画素2ないし画素9である。もちろん、異なった数や位置の近隣画素を用いることもできる。

【0042】ステップ505で、奇数番目の画素の平均と偶数番目の画素の平均とを求める。もう一度図11に戻ると、画素1, 3, 5, 7, 9の強度を平均して、奇数番目の画素の平均(AVG O)を求める。同様に、画素2, 4, 6, 8の強度を平均して、偶数番目の画素の平均(AVG E)を求める。すなわち、奇数番目の画素は、奇数番号が与えられた画素であり、偶数番目の画素は、偶数番号が与えられた画素である。

【0043】次に、ステップ507で、奇数番目の画素の平均が偶数番目の画素の平均よりも大きいかどうかを判定することにより、画素1をたたみ込んで、たたみ込み画像内のたたみ込み画素を生成する。奇数番目の画素の平均が偶数番目の画素の平均よりも大きい場合には、ステップ509で、たたみ込み画素は、奇数番目の画素の最小強度から偶数番目の画素の最大強度を引いた値に等しく設定される。そうでなければ、ステップ511で、たたみ込み画素が、偶数番目の画素の最小強度から奇数番目の画素の最大強度を引いた値に等しく設定される。

【0044】概念的には、近隣画素を、好ましい実施例

で用いられるソフトウェア・フィルタのようなフィルタで処理したと考えてもよい。この場合、フィルタを用いて、すべての奇数番目の画素が偶数番目の画素に比べて暗いか、あるいは、明るいかのいずれかであるような市松模様のパターンを検索する。そして、ステップ505で、奇数番目の画素の平均と偶数番目の画素の平均とを算出する。次に、ステップ507で、奇数番目の画素、即ち、正方形が明るい(即ち、高強度)あるいは暗い(即ち、低強度)市松模様のパターンを画素が反映しているかどうかの判定を行う。奇数番目の画素が明るい市松模様のパターンを反映している場合には、ステップ509で、所定の奇数番目の画素と所定の偶数番目の画素との差をたたみ込み画素に設定する。ここで、所定の奇数番目の画素は、奇数番目の画素の最小値であり、所定の偶数番目の画素は、偶数番目の画素の最大値である。ステップ511の処理は、ステップ509の逆になる。

【0045】このように、すべての奇数番目の画素がすべての偶数番目の画素よりも明るい場合には、ステップ509で設定される差は大きな値になる。従って、たたみ込み画素は、比較的明るい画素(即ち、高強度)になる。また、すべての偶数番目の画素がすべての奇数番目の画素よりも明るい場合にも、ステップ511で設定されるたたみ込み画素は、比較的明るい画素になる。一方、ステップ509あるいはステップ511で設定される差が非常に小さい(あるいは負である)場合には、たたみ込み画素は、比較的低い強度に設定される。たたみ込み画素が負の画素値を有する場合には、その値がゼロに設定されることが望ましい。簡単に言えば、フィルタが市松模様のパターンを見つけた場合には、たたみ込み画素が明るく、一方、フィルタが比較的ランダムなパターンを見つけた場合には、たたみ込み画素が暗くなる(この結果、所望のパターンではない「ノイズ」をフィルターで除去することができる)。

【0046】図9(b)の認識可能なパターン、即ち、グリッドパターンは、図10のソフトウェア・フィルタを用いることにより生成される。認識可能なパターンが生成される処理をより深く理解するために、図12(a)ないし図12(d)に、市松模様上でフィルタを動かして、たたみ込み画像内にグリッドパターンを生成する処理を示す。図12(a)の正方形530で示した走査画像内のパターン上でフィルタをたたみ込むことにより、市松模様のパターンを見つけて、たたみ込み画像内に明るい正方形を形成する。同様にして、図12(b)の正方形530で示したパターンをフィルタ処理することにより、たたみ込み画像内に明るい正方形を生成する。もちろん、図12(a)と図12(b)に示す正方形530内の市松模様のパターンは逆であるが、図10を参照して上述したように、いずれの場合でも、たたみ込み画像内に明るい正方形が生成される。図12(c)及び図12(d)でも、2つの明るい正方形が生

成される。この結果、図9(b)に示すように、2(2個の明るい正方形のグリッドパターンが生成される。

【0047】さらに、図10のソフトウェア・フィルタは、所望のパターンではない信号を除去する作用があるため、(グリッドパターンのような)認識可能なパターンを同定することが容易になる。たたみ込み画像内の認識可能なパターンを用いることにより、走査画像のアラ

	従来法
完全なアライメント	0%
1画素外れ	8%
2画素以上外れ	20%
1セル以上外れ	12%
アライン不能	60%

従来法では、(フィルタ処理されていない)走査画像を解析して、市松模様のパターンに含まれる明るい領域即ちスポットの位置決めを行った。この表に示すように、本発明の実施例では、走査画像のアライメントの精度を飛躍的に増大させることができた。

【0048】3. 改良グリッド・アライメント処理

走査画像のアライメントの精度を更に上げるために、改良画像アライメント処理を行うようにしてもよい。図13は、走査画像全体にわたって改良グリッド・アライメント処理を実行するためのフローチャートである。例えば、上述した処理を実行して走査画像をアラインした後、図13の処理を行って、アライメントの改良を実現するようにしてもよい。

【0049】ステップ551で、グリッド内に含まれるグリッド線上の画素強度を合計する。例えば、走査画像に含まれる市松模様のパターンにおいて、鉛直方向のグリッド強度を合計する。図14は、走査画像内のグリッド線を解析することにより、改良グリッド・アライメント処理を行う説明図である。図示するように、市松模様のパターン603に含まれる鉛直線601の画素強度を合計して、格納する。

【0050】次に、ステップ553で、解析すべきグリ

	従来法
完全なアライメント	0%
1画素外れ	8%
2画素以上外れ	20%
1セル以上外れ	12%
アライン不能	60%

従来法では、(フィルタ処理されていない)走査画像を解析して、市松模様のパターンに含まれる明るい領域即ちスポットの位置決めを行った。この表に示すように、本発明の実施例では、走査画像のアライメントの精度を飛躍的に増大させることができた。さらに、改良グリッド・アライメント処理によって、完全にアラインされた走査画像の割合を4%から6.4%に上げることができた。即ち、改良グリッド・アライメント処理を実行することにより、グリッド・アライメントの精度を有意に増

大化させることができる。再び図10に戻って、フィルタを用いて所定の画素をたたみ込み画素にたたみ込んだ後、ステップ513で、走査画像内に処理すべき画素が他にあるかどうかを判定する。以下に、本発明の実施例に従って、ハイブリッド形成されたチップの走査画像をアラインした例を示す。

フィルターを用いたたみ込み法

4%
9.6%
0%
0%
0%

ッド位置が他にあるかどうかの判定を行う。存在する場合には、ステップ555で、グリッドの位置を調節する。即ち、グリッドを左右に一画素分あるいはそれ以上移動して、その後、ステップ551で、グリッド線に沿って、強度の和を求める。グリッドに含まれるすべての位置の解析が完了した後、画素強度(即ち、ステップ551で算出した合計)が最小になるグリッド位置を選択する。他の位置におけるグリッド線の画素強度の方が低ければ、そのようにグリッドの調節を行う。このような改良処理は、特に、暗い領域あるいは線でセル同士が分離されている場合に、有効である。

【0051】図13の処理は、鉛直方向のグリッド線に関して説明したものであるが、同様のグリッド・アライメント処理を垂直方向に関しても実行することが望ましい。改良処理のために移動可能なグリッドの距離は限られている。例えば、セル・サイズの1/3の距離だけグリッドを移動することができる。

【0052】以下に、本発明の実施例に従って、改良グリッドアライメント処理を利用することにより、ハイブリッド形成されたチップの走査画像をアラインした例を示す。

改良グリッドアライメント法

6.4%
3.6%
0%
0%
0%

大化させることができる。

【0053】以上、本発明の好適な実施例を説明したが、本発明はこれらの実施例に何ら限定されるものではなく、さまざまに変形、変更、修正可能である。例えば、上述の実施例では、走査画像内に市松模様のパターンが存在するものについて説明した。が、本発明は、特定のパターンに限定されるものではなく、他のパターンにも同様に適用可能である。よって、実施例の説明は、本発明の範囲を何ら限定するものではない。

【図面の簡単な説明】

【図1】本発明の実施例のソフトウェアを実行するため
に用いることができるコンピュータシステムの例を示す
説明図である。

【図2】コンピュータシステムを示すシステムブロック
図である。

【図3】DNAやRNA等の生物学的試料のアレイを形
成し、解析するための全体システムを示す説明図であ
る。

【図4】チップを合成する処理を示す高水準フローチャ
ートである。

【図5】チップ上のプローブの結合を示す概念図であ
る。

【図6】チップをハイブリッド形成して、解析すること
により、実験結果を得るために処理を示すフローチャー
トである。

【図7】チップ上の個々のセルを示すように、走査画像
全体にわたってアライメントされたグリッドを示す説明
図である。

【図8】画像アライメント処理を示すフローチャートで
ある。

【図9】市松模様のパターンから形成されたグリッド・
パターンを含むたたみ込み画像を示す説明図である。

【図10】走査画像のたたみ込み処理を示すフローチャ
ートである。

【図11】近傍画素を解析することにより、たたみ込み
画像内にたたみ込み画素を生成する処理を示す説明図で
ある。

【図12】走査画像全体にわたってフィルターを移動す
ることにより、たたみ込み画像を形成する処理を示す説
明図である。

【図13】走査画像全体にわたって改良グリッド・アラ
イメント処理を実行するためのフローチャートである。

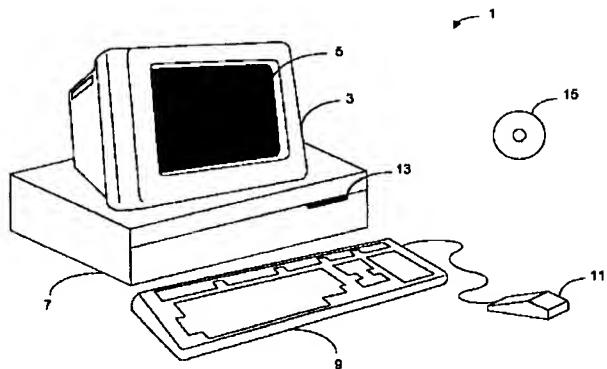
【図14】走査画像内のグリッド線を解析することによ
り、改良グリッド・アライメント処理を行う説明図であ
る。

【符号の説明】

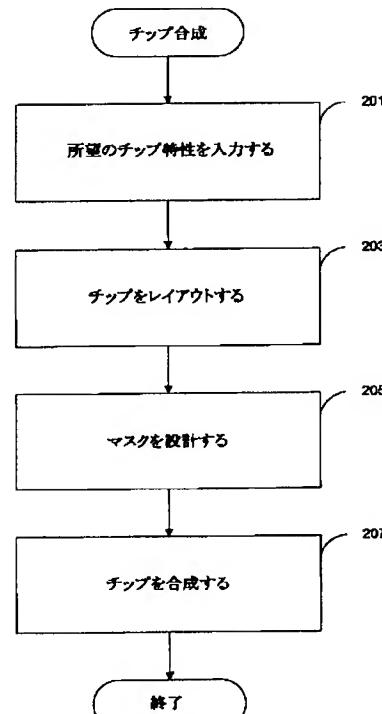
1…コンピュータシステム
3…ディスプレイ
5…スクリーン
7…キャビネット
9…キーボード
11…マウス
13…ROMドライブドライブ
15…ROM

5 1…中央処理装置
5 3…システムメモリ
5 5…固定記憶媒体
5 7…可動記憶媒体
5 9…ディスプレイアダプタ
6 1…サウンドカード
6 3…スピーカ
6 5…ネットワークインターフェース
6 7…矢印
1 0 0…コンピュータシステム
1 0 2…内部データベース
1 0 4…チップ設計コンピュータファイル
1 0 6…システム
1 0 8…ソフトウェア
1 1 0…マスク
1 1 2…合成システム
1 1 4…DNAプローブ
1 1 6…光源
1 1 8…フローセル
1 1 8…化学フローセル
1 1 9…コンピュータ
1 2 0…走査システム
1 2 2…デジタルコンピュータ
1 2 4…画像ファイル
1 2 6…解析システム
1 2 8…出力
2 5 1…チップ
2 5 3…試薬
2 5 7…チップ
2 6 1…デジタル画像
2 6 5…処理
2 6 9…セル強度ファイル
2 7 3…結果
3 0 1…チップ
3 0 1…走査画像
3 0 3…活性領域
3 0 5…パターン
3 0 7…グリッド
4 0 1…パターン
4 0 3…走査画像
4 5 1…パターン
4 5 3…画像
5 3 0…正方形
6 0 1…鉛直線
6 0 3…パターン

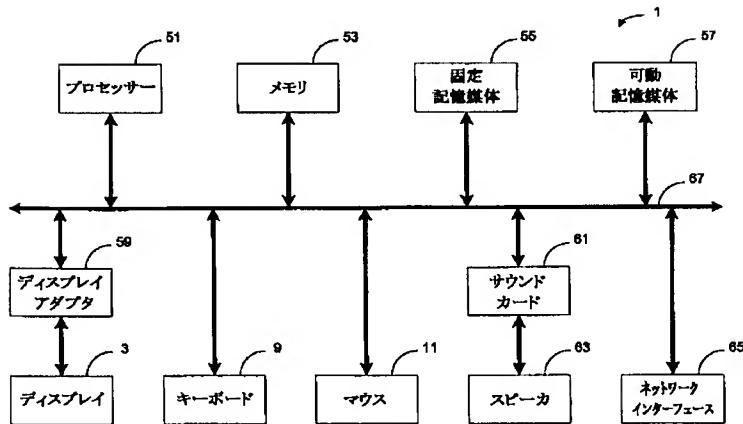
【図1】



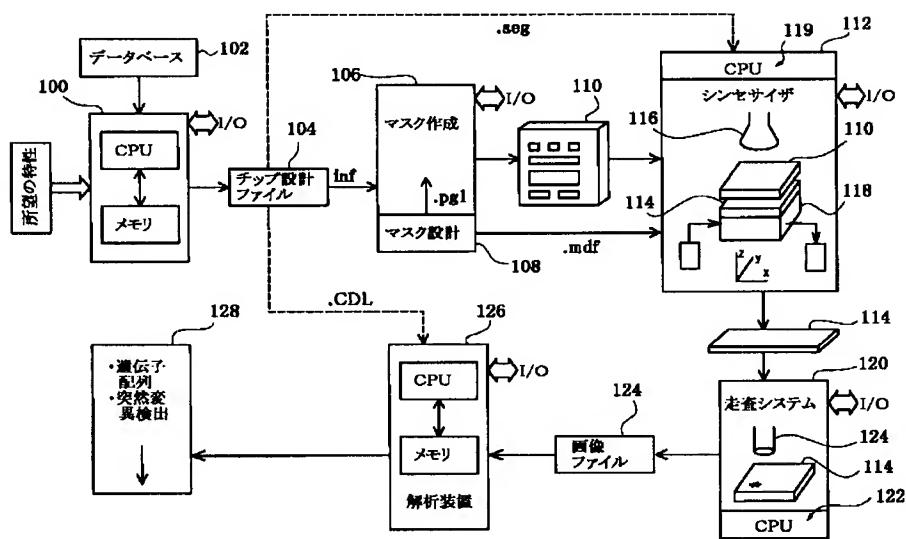
【図4】



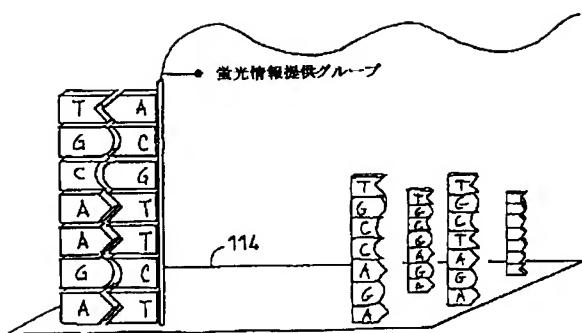
【図2】



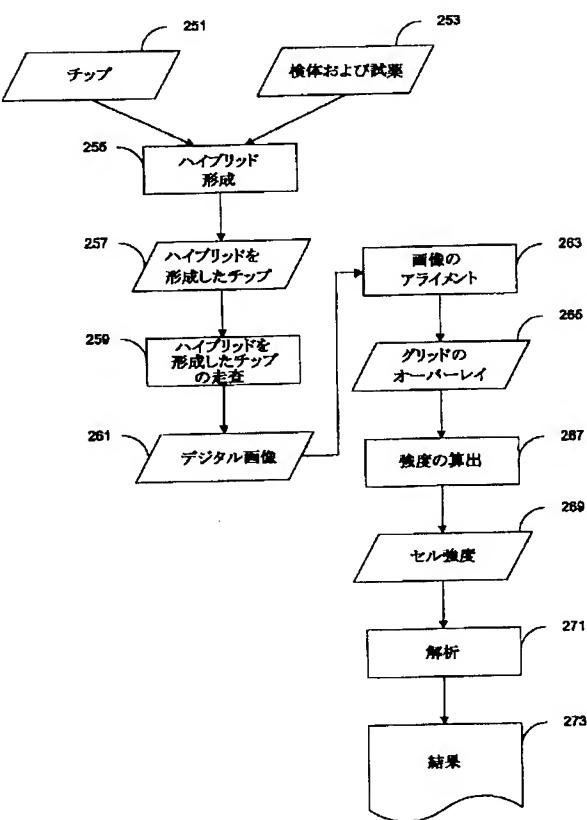
【図3】



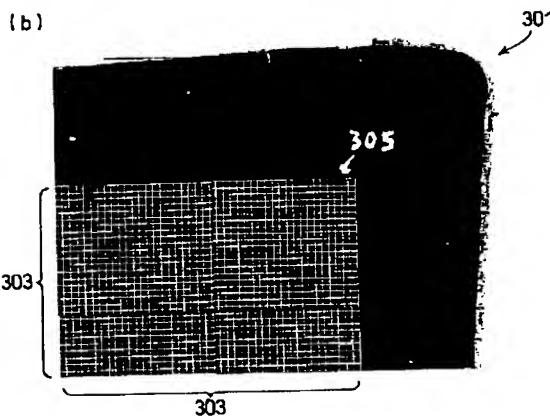
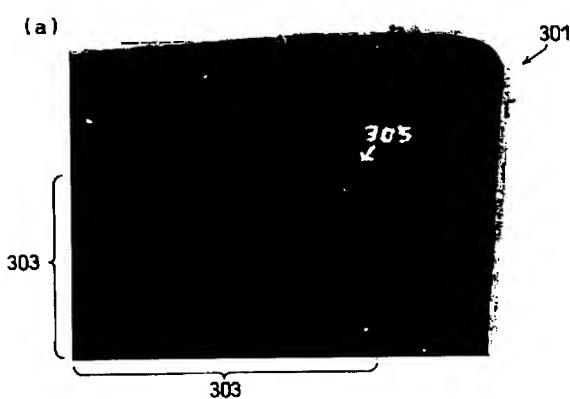
【図5】



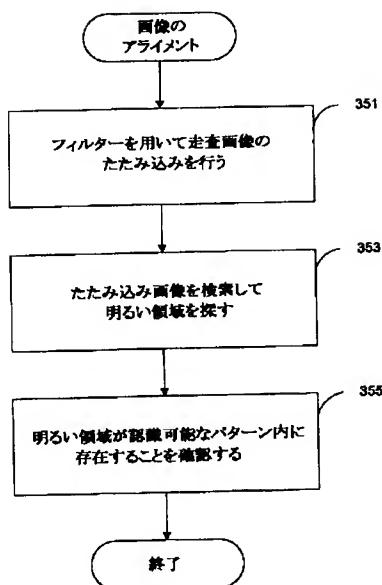
【図6】



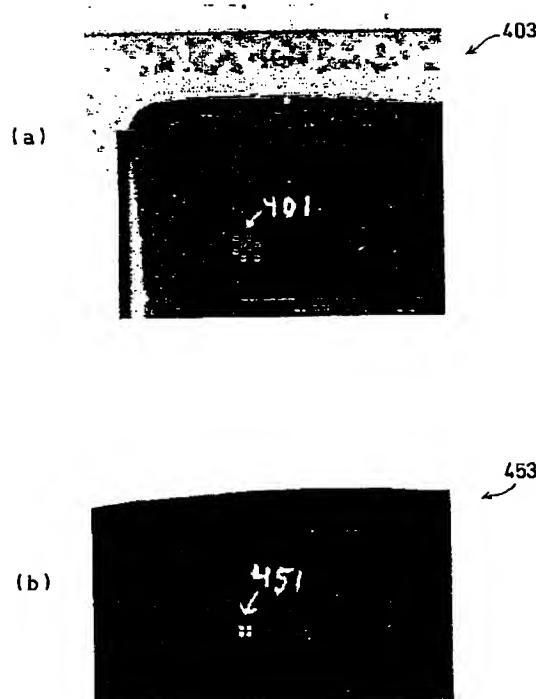
【図7】



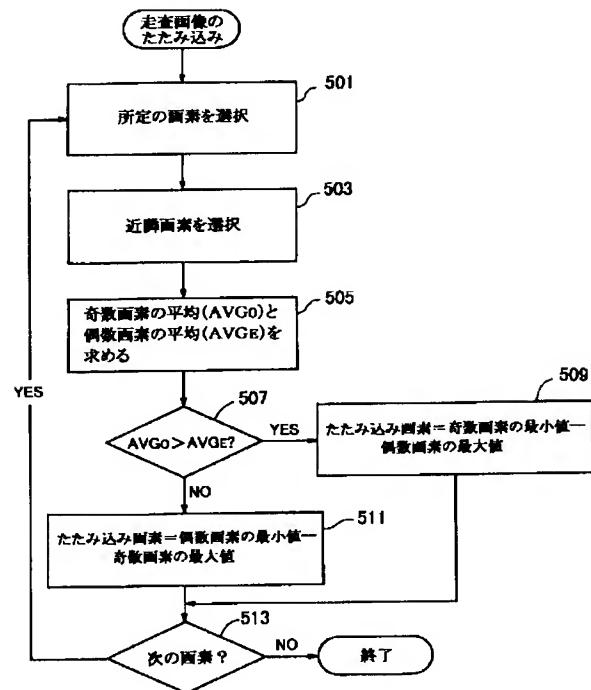
【図8】



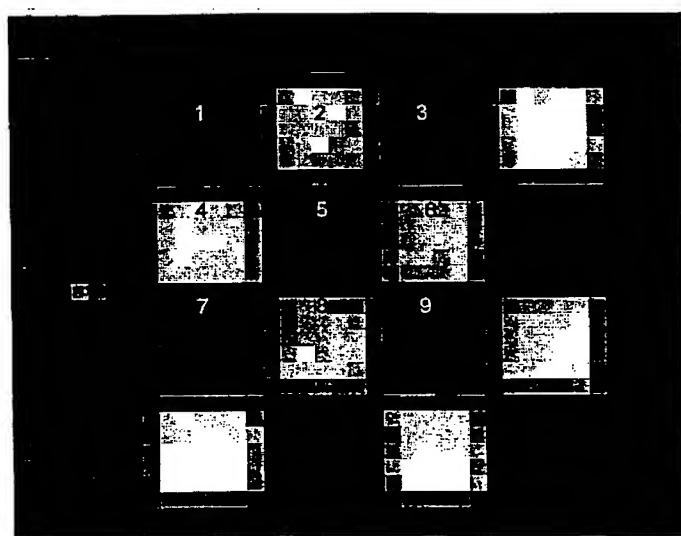
【図9】



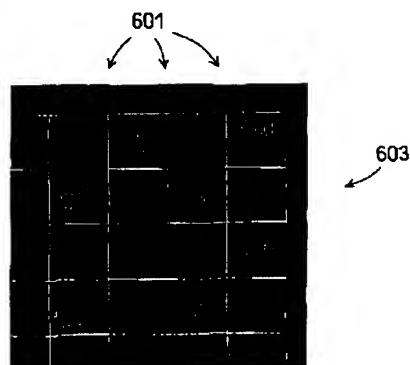
【図10】



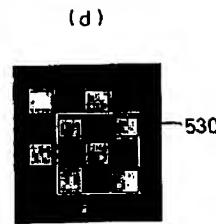
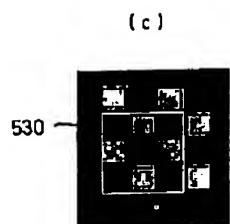
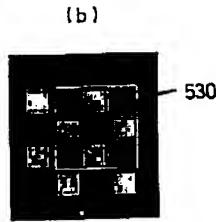
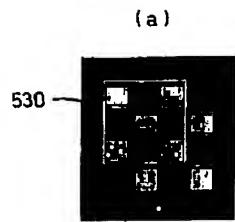
【図11】



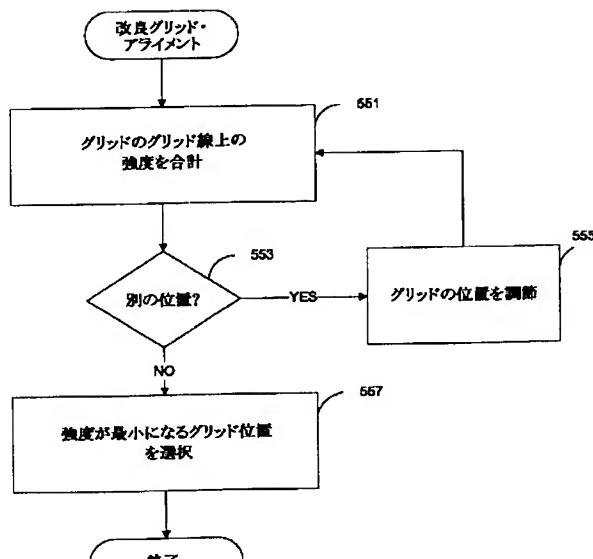
【図14】



【図12】



【図13】



フロントページの続き

(72)発明者 ピーター・フィーゴウスキイ
アメリカ合衆国 カリフォルニア州94024
ロス・アルトス, サウス・スプリング・
ロード, 952

(72)発明者 ダン・エム. . バーテル
アメリカ合衆国 カリフォルニア州94070
サン・カルロス, ブリッタン・アベニュー
-, 3321, アパートメント17

【外國語明細書】

1 Title of Invention

SCANNED IMAGE ALIGNMENT SYSTEMS AND METHODS

2 Claims

1. In a computer system, a method of aligning scanned images, comprising:
convolving a scanned image with a filter, the scanned image including a first pattern
that the filter will convolve into a second pattern in a convolved image; and
aligning the scanned image according to a position of the second pattern in the
convolved image.

2. The method of claim 1, wherein convolving a scanned image with a filter comprises
setting a convolved pixel to a difference between a selected odd pixel and a selected even
pixel of the first pattern.

3. The method of claim 2, wherein the selected odd pixel has the lowest intensity of the
odd pixels and the selected even pixel has the highest intensity of the even pixels, if the
average intensity of the odd pixels is greater than the average intensity of the even pixels.

4. The method of claim 2, wherein the selected odd pixel has the highest intensity of
the odd pixels and the selected even pixel has the lowest intensity of the even pixels, if the
average intensity of the odd pixels is not greater than the average intensity of the even
pixels.

5. The method of claim 1, wherein the first pattern is a checkerboard pattern.

6. The method of claim 1, wherein the second pattern is a grid pattern.

整理番号=PA57C683

ページ(2)

7. The method of claim 1, wherein aligning the scanned image comprises aligning a grid over the scanned image.

8. The method of claim 7, further comprising adjusting the position of the grid to minimize a sum of the intensities of pixels along a direction in the grid.

9. The method of claim 1, wherein the scanned image includes multiple copies of the first pattern.

10. The method of claim 9, wherein the scanned image is a rectangle with a copy of the first pattern near each corner.

11. A computer program product that aligns scanned images, comprising:
computer code that convolves a scanned image with a filter, the scanned image including a first pattern that the filter will convolve into a second pattern in a convolved image;
computer code that aligns the scanned image according to a position of the second pattern in the convolved image; and
a computer readable medium that stores the computer codes.

12. A method of aligning scanned images, comprising:
synthesizing a chip having attached nucleic acid sequences, the chip including a first pattern of nucleic acid sequences;
hybridizing labeled nucleic acid sequences to nucleic acid sequences on the chip;
scanning the hybridized chip to produce a scanned image;
convolving the scanned image with a filter, the filter convolving the first pattern into a second pattern in a convolved image; and
aligning the scanned image according to a position of the second pattern in the convolved image.

13. The method of claim 12, wherein convolving the scanned image with a filter comprises setting a convolved pixel to a difference between a selected odd pixel and a selected even pixel of the first pattern.

14. The method of claim 13, wherein the selected odd pixel has the lowest intensity of the odd pixels and the selected even pixel has the highest intensity of the even pixels, if the average intensity of the odd pixels is greater than the average intensity of the even pixels.

15. The method of claim 13, wherein the selected odd pixel has the highest intensity of the odd pixels and the selected even pixel has the lowest intensity of the even pixels, if the average intensity of the odd pixels is not greater than the average intensity of the even pixels.

16. The method of claim 12, wherein the first pattern is a checkerboard pattern.

17. The method of claim 16, wherein the labeled nucleic acid sequences include control nucleic acid sequences that hybridize to alternating squares in the checkerboard pattern.

18. The method of claim 12, wherein the second pattern is a grid pattern.

19. The method of claim 12, wherein aligning the scanned image comprises aligning a grid over the scanned image.

20. The method of claim 19, further comprising adjusting the position of the grid to minimize a sum of the intensities of pixels along a direction in the grid.

21. The method of claim 12, wherein the scanned image includes multiple copies of the first pattern.

22. The method of claim 21, wherein the scanned image is a rectangle with a copy of the first pattern near each corner.

23. A computer program product that aligns scanned images, comprising:
computer code that receives as input a scanned image of a chip having attached nucleic acid sequences to which labeled nucleic acid sequences are hybridized, the chip including a first pattern of nucleic acid sequences;

computer code that convolves the scanned image with a filter, the filter convolving the first pattern into a second pattern in a convolved image;

computer code that aligns the scanned image according to a position of the second pattern in the convolved image; and

a computer readable medium that stores the computer codes.

24. A chip, comprising:

a plurality of polymers attached to the chip in a first pattern so that the first pattern can be convolved into a second pattern to align the chip for scanning.

25. The chip of claim 24, wherein the first pattern is a checkerboard pattern.

26. The chip of claim 24, wherein the second pattern is a grid pattern.

27. The chip of claim 24, wherein the chip is a rectangle with the plurality of polymers attached to the chip in a first pattern near each corner.

28. The chip of claim 24, wherein the plurality of polymers are nucleic acid sequences.

3 Detailed Description of Invention

整理番号=P A 5 7 C 6 8 3

ページ (5)

The present invention relates to the field of image processing. More specifically, the present invention relates to computer systems for aligning grids on a scanned image of a chip including hybridized nucleic acid sequences.

Devices and computer systems for forming and using arrays of materials on a chip or substrate are known. For example, PCT applications WO92/10588 and 95/11995, both incorporated herein by reference for all purposes, describe techniques for sequencing or sequence checking nucleic acids and other materials. Arrays for performing these operations may be formed in arrays according to the methods of, for example, the pioneering techniques disclosed in U.S. Patent Nos. 5,445,934, 5,384,261 and 5,571,639, each incorporated herein by reference for all purposes.

According to one aspect of the techniques described therein, an array of nucleic acid probes is fabricated at known locations on a chip. A labeled nucleic acid is then brought into contact with the chip and a scanner generates an image file (also called a cell file) indicating the locations where the labeled nucleic acids are bound to the chip. Based upon the image file and identities of the probes at specific locations, it becomes possible to extract information such as the nucleotide or monomer sequence of DNA or RNA. Such systems have been used to form, for example, arrays of DNA that may be used to study and detect mutations relevant to genetic diseases, cancers, infectious diseases, HIV, and other genetic characteristics.

The VLSIPS™ technology provides methods of making very large arrays of oligonucleotide probes on very small chips. See U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092, each of which is incorporated by reference for all purposes. The oligonucleotide probes on the DNA probe array are used to detect complementary nucleic acid sequences in a sample nucleic acid of interest (the "target" nucleic acid).

For sequence checking applications, the chip may be tiled for a specific target nucleic acid sequence. As an example, the chip may contain probes that are perfectly complementary to the target sequence and probes that differ from the target sequence by a single base mismatch. For de novo sequencing applications, the chip may include all the

整理番号=PA57C683

ページ(6)

possible probes of a specific length. The probes are tiled on a chip in rows and columns of cells, where each cell includes multiple copies of a particular probe. Additionally, "blank" cells may be present on the chip which do not include any probes. As the blank cells contain no probes, labeled targets should not bind specifically to the chip in this area. Thus, a blank cell provides a measure of the background intensity.

In the scanned image file, a cell is typically represented by multiple pixels. Although a visual inspection of the scanned image file may be performed to identify the individual cells in the scanned image file. It would be desirable to utilize computer-implemented image processing techniques to align the scanned image file.

SUMMARY OF THE INVENTION

Embodiments of the present invention provide innovative techniques for aligning scanned images. A pattern is included in the scanned image so that when the image is convolved with a filter, a recognizable pattern is generated in the convolved image. The scanned image may then be aligned according to the position of the recognizable pattern in the convolved image. The filter may also act to remove or "filter out" the portions of the scanned image that do not correspond to the pattern in the scanned image. Several embodiments of the invention are described below.

In one embodiment, the invention provides a computer-implemented method of aligning scanned images. The scanned image is convolved with a filter. The scanned image includes a first pattern that the filter will convolve into a second pattern in the convolved image. The scanned image is then aligned according to the position of the second pattern in the convolved image. In a preferred embodiment, the first pattern may be a checkerboard pattern that is convolved into a grid pattern in the convolved image.

In another embodiment, the invention provides a method of aligning scanned images of chips with hybridized nucleic sequences. A chip having attached nucleic acid sequences (probes) is synthesized, with the chip including a first pattern of nucleic acid sequences. Labeled nucleic acid sequences are hybridized to nucleic acid sequences on the chip and the hybridized chip is scanned to produce a scanned image. The scanned image is convolved

整理番号=PA57C683

ページ(7)

with a filter that will convolve the first pattern into a second pattern in the convolved image. The scanned image is then aligned according to the position of the second pattern in the convolved image. In a preferred embodiment, the first pattern may be a checkerboard pattern that is generated by control nucleic acid sequences that hybridize to alternating squares in the checkerboard pattern.

Other features and advantages of the invention will become readily apparent upon review of the following detailed description in association with the accompanying drawings.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Overview

In the description that follows, the present invention will be described in reference to preferred embodiments that utilize VLSIPS™ technology for making very large arrays of oligonucleotide probes on chips. However, the invention is not limited to images produced in this fashion and may be advantageously applied other hybridization technologies or images in other technology areas. Therefore, the description of the embodiments that follows for purposes of illustration and not limitation.

Fig. 1 illustrates an example of a computer system that may be used to execute the software of an embodiment of the invention. Fig. 1 shows a computer system 1 that includes a display 3, screen 5, cabinet 7, keyboard 9, and mouse 11. Mouse 11 may have one or more buttons for interacting with a graphical user interface. Cabinet 7 houses a CD-ROM drive 13, system memory and a hard drive (see Fig. 2) which may be utilized to store and retrieve software programs incorporating computer code that implements the invention, data for use with the invention, and the like. Although a CD-ROM 15 is shown as an exemplary computer readable storage medium, other computer readable storage media including floppy disk, tape, flash memory, system memory, and hard drive may be utilized. Additionally, a data signal embodied in a carrier wave (e.g., in a network including the Internet) may be the computer readable storage medium.

Fig. 2 shows a system block diagram of computer system 1 used to execute the software of an embodiment of the invention. As in Fig. 1, computer system 1 includes

整理番号=P A 5 7 C 6 8 3

ページ (8)

monitor 3 and keyboard 9, and mouse 11. Computer system 1 further includes subsystems such as a central processor 51, system memory 53, fixed storage 55 (e.g., hard drive), removable storage 57 (e.g., CD-ROM drive), display adapter 59, sound card 61, speakers 63, and network interface 65. Other computer systems suitable for use with the invention may include additional or fewer subsystems. For example, another computer system could include more than one processor 51 (i.e., a multi-processor system) or a cache memory.

The system bus architecture of computer system 1 is represented by arrows 67. However, these arrows are illustrative of any interconnection scheme serving to link the subsystems. For example, a local bus could be utilized to connect the central processor to the system memory and display adapter. Computer system 1 shown in Fig. 2 is but an example of a computer system suitable for use with the invention. Other computer architectures having different configurations of subsystems may also be utilized.

The present invention provides methods of aligning scanned images or image files of hybridized chips including nucleic acid probes. In a representative embodiment, the scanned image files include fluorescence data from a biological array, but the files may also represent other data such as radioactive intensity, light scattering, refractive index, conductivity, electroluminescence, or large molecule detection data. Therefore, the present invention is not limited to analyzing fluorescence measurements of hybridization but may be readily utilized to analyze other measurements of hybridization.

For purposes of illustration, the present invention is described as being part of a computer system that designs a chip mask, synthesizes the probes on the chip, labels the nucleic acids, and scans the hybridized nucleic acid probes. Such a system is fully described in U.S. Patent No. 5,571,639 that has been incorporated by reference for all purposes. However, the present invention may be used separately from the overall system for analyzing data generated by such systems.

Fig. 3 illustrates a computerized system for forming and analyzing arrays of biological materials such as RNA or DNA. A computer 100 is used to design arrays of biological polymers such as RNA and DNA. The computer 100 may be, for example, an appropriately programmed Sun Workstation or personal computer or workstation, such as

整理番号 = P A 5 7 C 6 8 3

ページ (9)

an IBM PC equivalent, including appropriate memory and a CPU as shown in Figs. 1 and 2.

The computer system 100 obtains inputs from a user regarding characteristics of a gene of interest, and other inputs regarding the desired features of the array. Optionally, the computer system may obtain information regarding a specific genetic sequence of interest from an external or internal database 102 such as GenBank. The output of the computer system 100 is a set of chip design computer files 104 in the form of, for example, a switch matrix, as described in PCT application WO 92/10092, and other associated computer files.

The chip design files are provided to a system 106 that designs the lithographic masks used in the fabrication of arrays of molecules such as DNA. The system or process 106 may include the hardware necessary to manufacture masks 110 and also the necessary computer hardware and software 108 necessary to lay the mask patterns out on the mask in an efficient manner. As with the other features in Fig. 3, such equipment may or may not be located at the same physical site but is shown together for ease of illustration in Fig. 3. The system 106 generates masks 110 or other synthesis patterns such as chrome-on-glass masks for use in the fabrication of polymer arrays.

The masks 110, as well as selected information relating to the design of the chips from system 100, are used in a synthesis system 112. Synthesis system 112 includes the necessary hardware and software used to fabricate arrays of polymers on a substrate or chip 114. For example, synthesizer 112 includes a light source 116 and a chemical flow cell 118 on which the substrate or chip 114 is placed. Mask 110 is placed between the light source and the substrate/chip, and the two are translated relative to each other at appropriate times for deprotection of selected regions of the chip. Selected chemical reagents are directed through flow cell 118 for coupling to deprotected regions, as well as for washing and other operations. All operations are preferably directed by an appropriately programmed computer 119, which may or may not be the same computer as the computer(s) used in mask design and mask making.

The substrates fabricated by synthesis system 112 are optionally diced into smaller chips and exposed to marked targets. The targets may or may not be complementary to one or more of the molecules on the substrate. The targets are marked with a label such as a

整理番号=PA57C683

ページ(10)

fluorescein label (indicated by an asterisk in Fig. 3) and placed in scanning system 120. Scanning system 120 again operates under the direction of an appropriately programmed digital computer 122, which also may or may not be the same computer as the computers used in synthesis, mask making, and mask design. The scanner 120 includes a detection device 124 such as a confocal microscope or CCD (charge-coupled device) that is used to detect the location where labeled target (*) has bound to the substrate. The output of scanner 120 is an image file(s) 124 indicating, in the case of fluorescein labeled target, the fluorescence intensity (photon counts or other related measurements, such as voltage) as a function of position on the substrate. Since higher photon counts will be observed where the labeled target has bound more strongly to the array of polymers (e.g., DNA probes on the substrate), and since the monomer sequence of the polymers on the substrate is known as a function of position, it becomes possible to determine the sequence(s) of polymer(s) on the substrate that are complementary to the target.

The image file 124 is provided as input to an analysis system 126 that incorporates the scanned image alignment techniques of the present invention. Again, the analysis system may be any one of a wide variety of computer system(s), but in a preferred embodiment the analysis system is based on a WINDOWS NT workstation or equivalent. The analysis system may analyze the image file(s) to generate appropriate output 128, such as the identity of specific mutations in a target such as DNA or RNA.

Fig. 4 is a high level flowchart of a process of synthesizing a chip. At a step 201, the desired chip characteristics are input to the chip synthesis system. The chip characteristics may include (such as sequence checking systems) the genetic sequence(s) or targets that would be of interest. The sequences of interest may, for example, identify a virus, microorganism or individual. Additionally, the sequence of interest may provide information about genetic diseases, cancers or infectious diseases. Sequence selection may be provided via manual input of text files or may be from external sources such as GenBank. In a preferred embodiment that performs de novo sequencing of target nucleic acids, this step is not necessary as the chip includes all the possible n-mer probes (where n represents the length of the nucleic acid probe).

整理番号 = P A 5 7 C 6 8 3

ページ (11)

For de novo sequencing, a chip may be synthesized to include cells containing all the possible probes of a specific length. For example, a chip may be synthesized that includes all the possible 8-mer DNA probes. Such a chip would have 65,536 cells (4^8), with each cell corresponding to a particular probe. A chip may also include other probes including all the probes of other lengths.

At a step 203 the system determines which probes would be desirable on the chip, and provides an appropriate "layout" on the chip for the probes. The layout implements desired characteristics such as an arrangement on the chip that permits "reading" of genetic sequence and/or minimization of edge effects, ease of synthesis, and the like.

The masks for the chip synthesis are designed at a step 205. The masks are designed according to the desired chip characteristics and layout. At a step 207, the system synthesizes the DNA or other polymer chips. Software controls, among other things, the relative translation of the substrate and mask, the flow of the desired reagents through a flow cell, the synthesis temperature of the flow cell, and other parameters.

Fig. 5 illustrates the binding of a particular target DNA to an array of DNA probes 114. As shown in this simple example, the following probes are formed in the array:

3'-AGAACGT

AGACCGT

AGAGCGT

AGATCGT

•

•

•

As shown, when the fluorescein-labeled (or otherwise marked) target 5'-TCTTGCA is exposed to the array, it is complementary only to the probe 3'-AGAACGT, and fluorescein will be primarily found on the surface of the chip where 3'-AGAACGT is located. The chip contains cells that include multiple copies of a particular probe. Thus, the image file

整理番号=P A 5 7 C 6 8 3

ページ (12)

will contain fluorescence intensities, one for each probe (or cell). By analyzing the fluorescence intensities associated with a specific probe, it becomes possible to extract sequence information from such arrays using the methods of the invention disclosed herein.

For ease of reference, one may call bases by assigning the bases the following codes:

Code	Group	Meaning
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T(U)	Thymine (Uracil)
M	A or C	aMino
R	A or G	puRine
W	A or T(U)	Weak interaction (2 H bonds)
Y	C or T(U)	pYrimidine
S	C or G	Strong interaction (3 H bonds)
K	G or T(U)	Keto
V	A, C or G	not T(U)
H	A, C or T(U)	not G
D	A, G or T(U)	not C
B	C, G or T(U)	not A
N	A, C, G, or T(U)	Insufficient intensity to call
X	A, C, G, or T(U)	Insufficient discrimination to call

Most of the codes conform to the IUPAC standard. However, code N has been redefined and code X has been added.

整理番号=P A 5 7 C 6 8 3

ページ (13)

Scanned Image Alignment

Before the scanned image alignment of the invention are discussed, it may be helpful to provide an overview of the overall process in one embodiment. Fig. 6 illustrates a flowchart of a process of how a chip is hybridized and analyzed to produce experimental results. A chip 251 having attached nucleic acid sequences (or probes) is combined with a sample nucleic acid sequence (e.g., labeled fragments of the sample) and reagents in a hybridization step 255. The hybridization step produces a hybridized chip 257.

The hybridized chip is scanned at a step 259. For example, the hybridized chip may be laser scanned to detect where fluoresccin-labeled sample fragments have hybridized to the chip. Numerous techniques may be utilized to label the sample fragments and the scanning process will typically be performed according to the type of label utilized. The scanning step produces a digital image of the chip.

In preferred embodiments, the scanned image of the chip includes varying fluorescent intensities that correspond to the hybridization intensity or affinity of the sample to the probes in a cell. In order to achieve more accurate results, it is beneficial to identify the pixels that belong to each cell on the chip. At an image alignment step 263, the scanned image is aligned so that the pixels that correspond to each cell can be identified. Optionally, the image alignment step includes the alignment of a grid over the scanned image (see Fig. 7B).

At a step 267, the analysis system analyzes the scanned image to calculate the relative hybridization intensities for each cell of interest on the chip. For example, the hybridization intensity for a cell, and therefore the relative hybridization affinity between the probe of the cell and the sample sequence, may be calculated as the mean of the pixel values within the cell. The pixel values may correspond to photon counts from the labeled hybridized sample fragments.

The cell intensities may be stored as a cell intensity file 269. In preferred embodiments, the cell intensity file includes a list of cell intensities for the cells. At an analysis step 271, the analysis system may analyze the cell intensity file and chip characteristics to generate results 273. The chip characteristics may be utilized to identify

整理番号=PA57C683

ページ (14)

the probes that have been synthesized at each cell on the chip. By analyzing both the sequence of the probes and their hybridization intensities from the cell intensity file, the system is able to extract sequence information such as the location of mutations, deletions or insertions, or the sequence of the sample nucleic acid. Accordingly, the results may include sequence information, graphs of the hybridization intensities of probe(s), graphs of the differences between sequences, and the like. See U.S. Patent Application No. 08/327,525, which is hereby incorporated by reference for all purposes.

In order to align the scanned image, the invention provides a pattern in the scanned image that will be convolved into a recognizable pattern. In preferred embodiments, the pattern in the scanned image is a checkerboard pattern that is generated by synthesizing alternating cells that include probes that are complementary to a control nucleic acid sequence. The control nucleic acid sequence may be a known sequence that is labeled and hybridized to the chip for the purpose of aligning the scanned image. Additionally, the brightness of the cells complementary to the control nucleic acid sequence may be utilized as a baseline or for comparison to other intensities.

As an example, Fig. 7A shows a checkerboard pattern in a hybridized chip. A scanned image 301 of a hybridized chip includes an active area 303 where the probes were synthesized. At the corner of the active area is a pattern 305 that is a checkerboard pattern. Typically, the pattern appears at each corner of the active area of the scanned image. Although the pattern is shown as being a checkerboard pattern, in other embodiments the pattern is a circle, square, plus sign, or any other pattern.

With regard to Fig. 6, it was stated that a grid may optionally be placed over the scanned image to show or delineate the individual cells of the chip. Fig. 7B shows a grid that has been aligned over the scanned image of Fig. 7A to show the individual cells of the chip. As shown, a grid 307 has been placed over active area 303 of hybridized chip 301.

Fig. 8 illustrates a flowchart of a process of image alignment. The flowchart shows detail for step 263 of Fig. 6. At a step 351, the scanned image is convolved with a filter. The filter is typically a software filter that convolves the scanned image into a convolved image. When the scanned image is convolved, a pattern in the scanned image is convolved

整理番号 = P A 5 7 C 6 8 3

ページ (15)

into a recognizable pattern. The position of the recognizable pattern in the convolved image may be utilized to align the scanned image, such as by placing a grid over the image.

At a step 353, the convolved image is searched for bright areas. When the scanned image is convolved, the pattern(s) in the scanned image will be convolved into a recognizable pattern or patterns of bright areas. Accordingly, once bright areas are identified in the convolved image, the system confirms that the bright areas are in the expected recognizable pattern (e.g., a grid pattern) at a step 355.

In order to better understand what is meant by the different patterns, Fig. 9A shows a checkerboard pattern 401 in a scanned image 403. Fig. 9B shows a recognizable pattern 451 in convolved image 453. The convolved image was generated from the scanned image of Fig. 9A. As shown, recognizable pattern 451 in this embodiment is a grid pattern that was generated by the checkerboard pattern when it was convolved with a filter. Additionally, it should be noted that the filter acted to remove the other pixel intensities so that the convolved image only includes the recognizable pattern. By removing pixel intensities that are not part of the pattern in the scanned image, it is easier to align the scanned image.

Fig. 10 illustrates a flowchart of a process of convolving the scanned image. The flowchart illustrates a process that may be performed at step 351 of Fig. 8. At a step 501, a pixel is selected. For simplicity, we will assume that the process selects pixels of the scanned image from left to right and top to bottom. Of course, the order that the pixels are analyzed may be varied.

Once a pixel selected, neighbor pixels may then be selected at a step 503. By neighbor pixels, it is meant pixels that the pixels are near, but not necessarily adjacent to a pixel. For example, Fig. 11 shows neighbor pixels that may be analyzed to produce a convolved pixel in a convolved image. As shown in Fig. 11, there are 9 pixels labeled 1-9. In a preferred embodiment, pixel 1 is the pixel retrieved at step 501 and the neighbor pixels retrieved at step 503 are pixels 2-9. Of course, any number or location of different neighbor pixels may be utilized.

整理番号 = P A 5 7 C 6 8 3

ページ (16)

At a step 505, the average of the odd pixels and the average of the even pixels is determined. Referring again to Fig. 11, the intensities of pixels 1, 3, 5, 7, and 9 may be averaged to produce the average of the odd pixels (AVG_O). Similarly, the intensities of pixels 2, 4, 6, and 8 may be averaged to produce the average of the even pixels (AVG_E). Thus, the odd pixels may be pixels that have an odd number designation and the even pixels may be pixels that have an even number designation.

Pixel 1 is convolved into a convolved pixel in a convolved image by determining if the average of the odd pixels is greater than the average of the even pixels at a step 507. If the average of the odd pixels is greater, the convolved pixel is set equal to the intensity of the minimum of the odd pixels minus the intensity of the maximum of the even pixels at a step 509. Otherwise, the convolved pixel is set equal to the intensity of the minimum of the even pixels minus the intensity of the maximum of the odd pixels at a step 511.

Conceptually, the neighbor pixels may be thought of as being filtered, such as by a software filter in preferred embodiments. With the filter, the system is searching for a checkerboard pattern where all the odd pixels are either darker or lighter than the even pixels. Accordingly, averages of the odd and even pixels are calculated at step 505. Step 507 acts to determine if the pixels likely reflect a checkerboard pattern where the odd pixels, and therefore squares, are light (e.g., high intensity) or dark (e.g., low intensity). If the odd pixels likely reflect a checkerboard pattern where the odd pixels are light, step 509 sets the convolved pixel to the difference between selected odd and even pixels, where the selected odd pixel is the minimum of the odd pixels and the selected even pixel is the maximum of the even pixels. Step 511 is similar but reversed.

Therefore, at step 509, if all the odd pixels are much brighter than all the even pixels, the difference will be a larger value. Hence, the convolved pixel will be relatively bright (e.g., high intensity). The convolved pixel will also be relatively bright if all the even pixels are much brighter than all the odd pixels at step 511. However, if the difference at step 509 or 511 is very small (or negative), the convolved pixel will be set to a relatively dark intensity. Convolved pixels with negative pixel values may be set to a zero in preferred embodiments. In short, if the filter finds a checkerboard pattern, the convolved pixel will be

整理番号=PA57C683

ページ(17)

bright and if the filter finds a relatively random pattern, the convolved pixel will be dark (thus, filtering out "noise" that is not the desired pattern).

The recognizable pattern in Fig. 9B, which is a grid pattern, was generated by the software filter of Fig. 10. In order to better see how the recognizable pattern was generated, Figs. 12A-D show how the filter may be moved over the checkerboard to produce a grid pattern in the convolved image. As the filter is convolved over the pattern in the scanned image shown in square 530 in Fig. 12A, a bright square will be generated in the convolved image since a checkerboard pattern will be found. Similarly, a bright square will be generated in the convolved image when the filter is over the pattern in square 530 of Fig. 12B. Of course, the checkerboard patterns in square 530 of Figs. 12A and 12B are reversed, but both will produce a bright square in the convolved image as described above in reference to Fig. 10. Figs. 12C and 12D will also produce two bright squares. Therefore, a 2x2 bright square grid pattern is generated as shown in Fig. 9B.

Additionally, as the software filter of Fig. 10 acts to filter out signals that are not the desired pattern, the recognizable pattern (e.g., a grid pattern) is easier to identify. The recognizable patterns in the convolved image are utilized to align the scanned image. Returning now to Fig. 10, after a selected pixel is convolved into a convolved pixel by the filter, it is determined if there is another pixel to process in the scanned image at a step 513.

The following shows how well an embodiment of the invention aligned scanned images of hybridized chips:

	Previous method	With filter convolution
Perfect alignment	0%	4%
1 pixel off	8%	96%
2 or more pixels off	20%	0%
1 or more cells off	12%	0%
unable to align	60%	0%

The previous method was to analyze the scanned image (unfiltered) to locate bright areas or spots in a checkerboard pattern. As shown, an embodiment of the invention was able to dramatically increase the accuracy of scanned image alignment.

Refined Grid Alignment

In preferred embodiments, refined image alignment may be performed to further increase the accuracy of the scanned image alignment. Fig. 13 illustrates a flowchart of a process of refining grid alignment over a scanned image. Thus, for example, once the above-described process has been performed to align the scanned image, the process in Fig. 13 may be utilized to refine the alignment.

At a step 551, pixel intensities on grid lines in the grid are summed. For example, the intensities of the grid in a vertical direction in the checkerboard pattern in the scanned image may be summed. Fig. 14 shows the grid lines in the scanned image that may be analyzed to refine the grid alignment. As shown, the pixel intensities of vertical lines 601 of a checkerboard pattern 603 may be summed and stored.

Then, at a step 553, the system may determine if there are more positions of the grid to analyze. If there are, the position of the grid may be adjusted at a step 555. Therefore, the grid may be moved left and right by one or more pixels before the intensities are summed along grid lines at step 551. Once all the positions of the grid have been analyzed, the system selects a grid position where pixel intensities (e.g., the sum calculated at step 551) are at a minimum. Therefore, if the pixel intensities for grid lines are lower at another position, the grid is adjusted accordingly. This refinement will work well if the cells are typically separated by a darker area or line.

Although the process in Fig. 13 was described for grid lines in the vertical direction, preferred embodiments also perform the same grid alignment for the horizontal direction. The distance that the grid is able to be moved for refinement may be limited. For example, the grid may be limited to movement of one-third a cell size.

The following shows how well an embodiment of the invention aligned scanned images of hybridized chips utilizing the refined grid alignment:

	Previous method	With refined grid alignment
Perfect alignment	0%	64%
1 pixel off	8%	36%

整理番号 = P A 5 7 C 6 8 3

ページ (19)

2 or more pixels off	20%	0%
1 or more cells off	12%	0%
unable to align	60%	0%

Once again, the previous method was to analyze the scanned image (unfiltered) to locate bright areas or spots in a checkerboard pattern. As shown, an embodiment of the invention was able to dramatically increase the accuracy of scanned image alignment. Furthermore, refining grid alignment increased the percentage of scanned images that were perfectly aligned with the invention from 4% to 64%. Therefore, performing a refinement of grid alignment can significantly increase the accuracy of the grid alignment.

Conclusion

While the above is a complete description of preferred embodiments of the invention, various alternatives, modifications, and equivalents may be used. It should be evident that the invention is equally applicable by making appropriate modifications to the embodiments described above. For example, the invention has been described in reference to a checkerboard pattern in the scanned image. However, the invention is not limited to any one pattern and may be advantageously applied to other patterns including those described herein. Therefore, the above description should not be taken as limiting the scope of the invention that is defined by the metes and bounds of the appended claims along with their full scope of equivalents.

4 Brief Description of Drawings

Fig. 1 illustrates an example of a computer system that may be utilized to execute the software of an embodiment of the invention.

Fig. 2 illustrates a system block diagram of the computer system of Fig. 1.

Fig. 3 illustrates an overall system for forming and analyzing arrays of biological materials such as DNA or RNA.

Fig. 4 is a high level flowchart of a process of synthesizing a chip.

Fig. 5 illustrates conceptually the binding of probes on chips.

整理番号= P A 5 7 C 6 8 3

ページ (20)

Fig. 6 illustrates a flowchart of how a chip is hybridized and analyzed to produce experimental results.

Fig. 7A shows a checkerboard pattern in a scanned image and Fig. 7B shows a grid that has been aligned over the scanned image to show the individual cells on the chip.

Fig. 8 illustrates a flowchart of a process of image alignment.

Fig. 9A shows a checkerboard pattern in a scanned image and Fig. 9B shows a convolved image of Fig. 9A with a grid pattern that was generated by the checkerboard pattern.

Fig. 10 illustrates a flowchart of a process of convolving the scanned image.

Fig. 11 shows neighbor pixels that may be analyzed to produce a convolved pixel in the convolved image.

Figs. 12A-12D show how the filter may be moved over the scanned image to produce the convolved image.

Fig. 13 illustrates a flowchart of a process of refining the grid alignment over the scanned image.

Fig. 14 shows the grid lines in the scanned image that may be analyzed to refine the grid alignment.

(35) 2000-69998 (P2000-6<A)

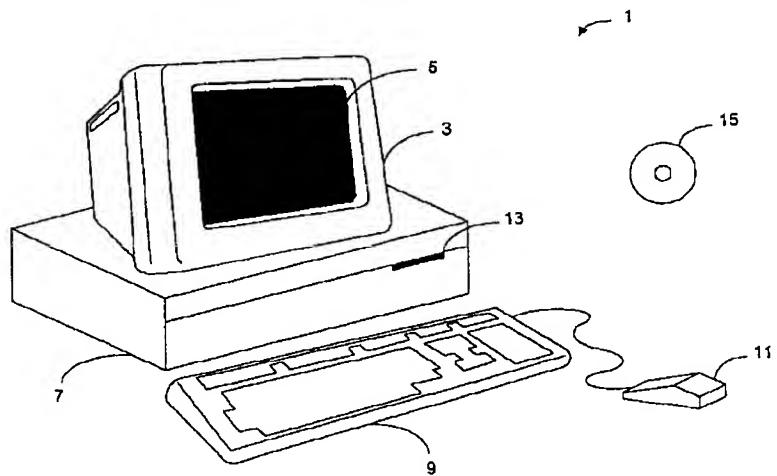


FIG. 1

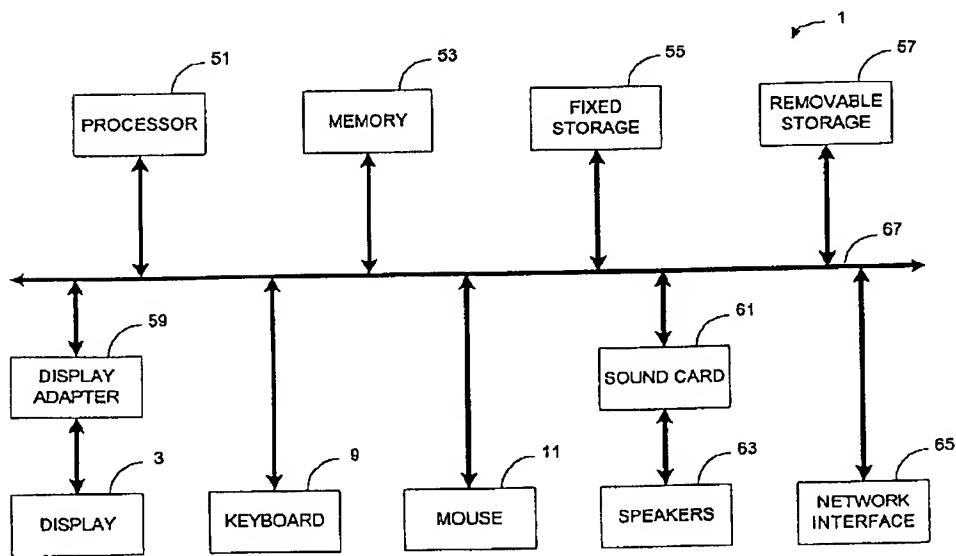


FIG. 2

整理番号=PA57C683

2-2 (2)

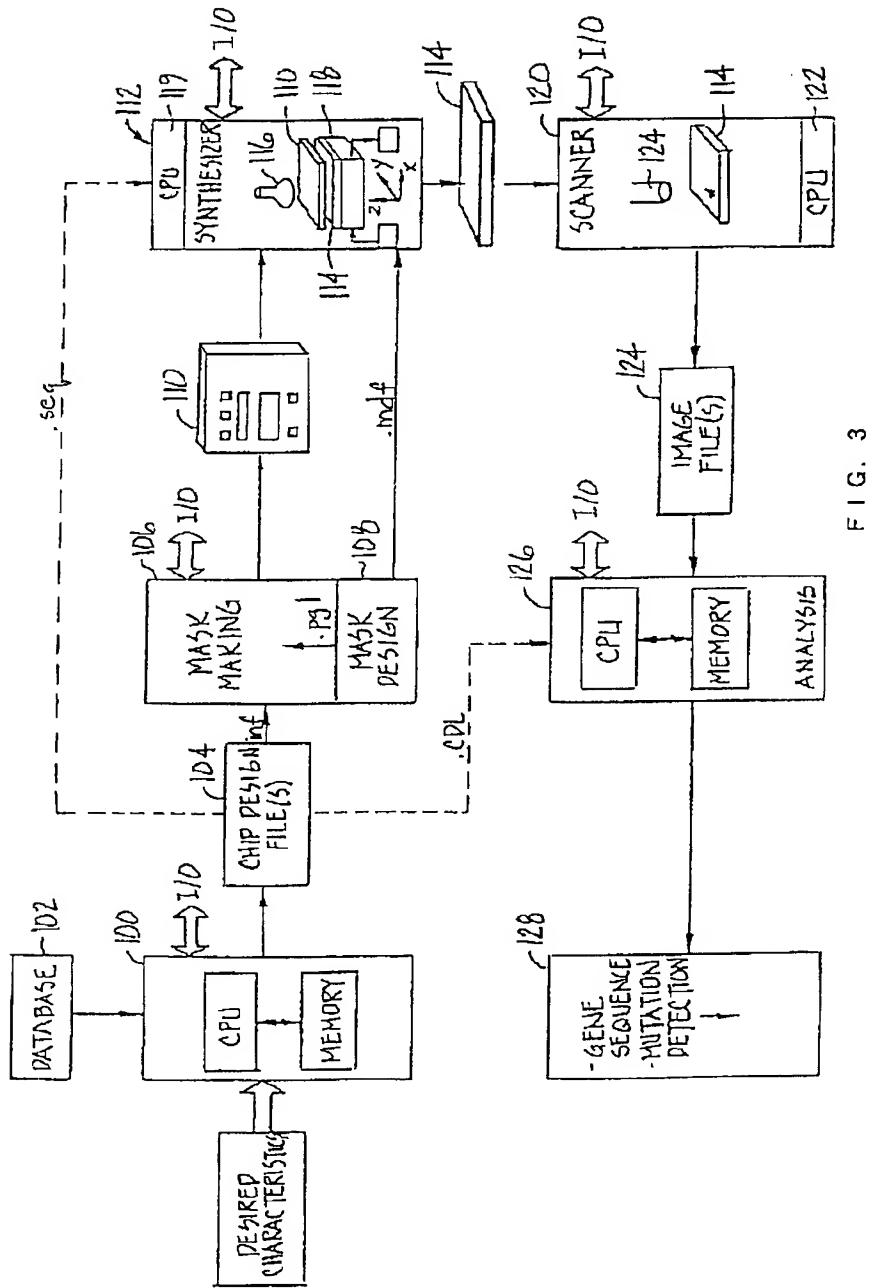


FIG. 3

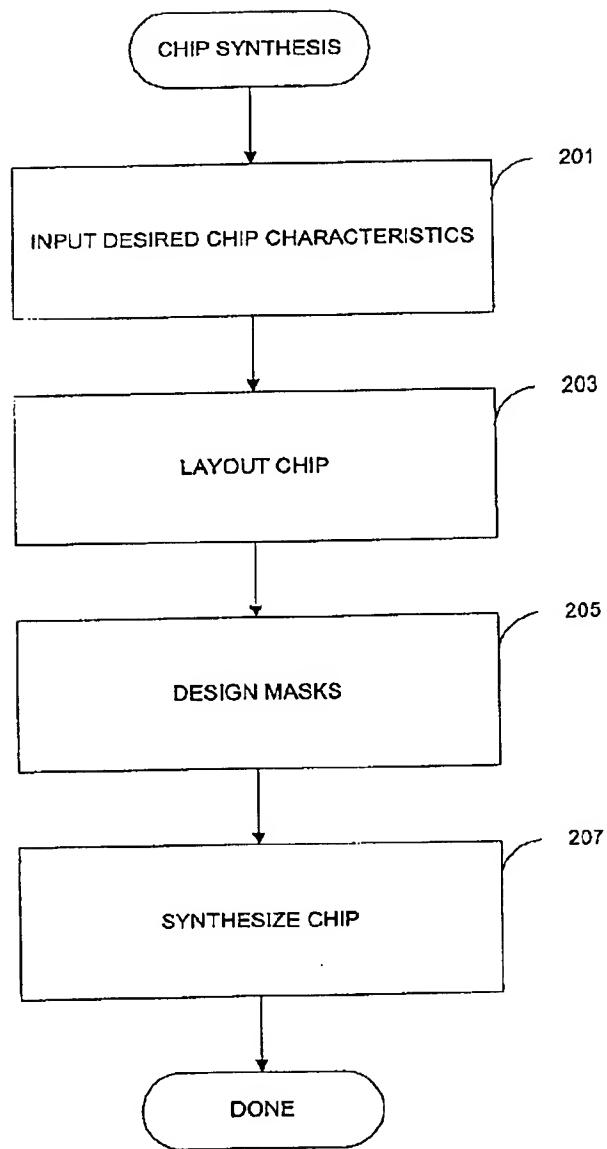


FIG. 4

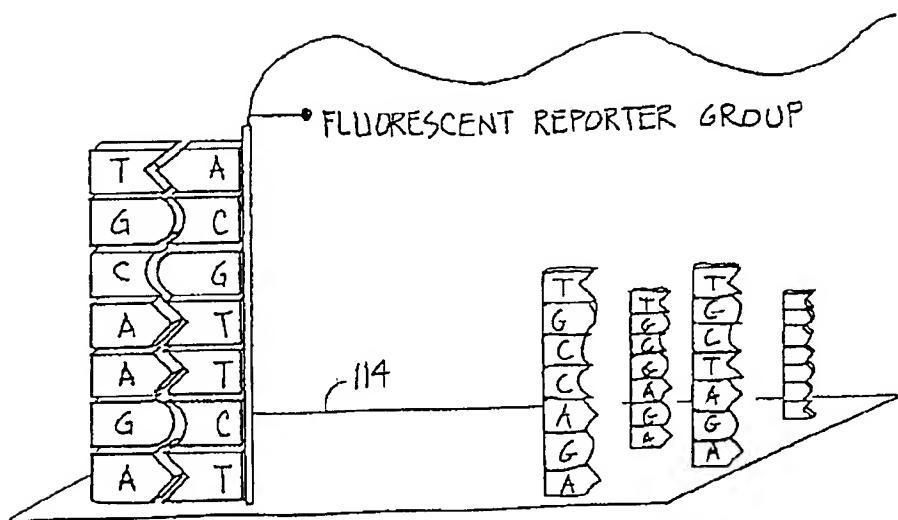


FIG. 5

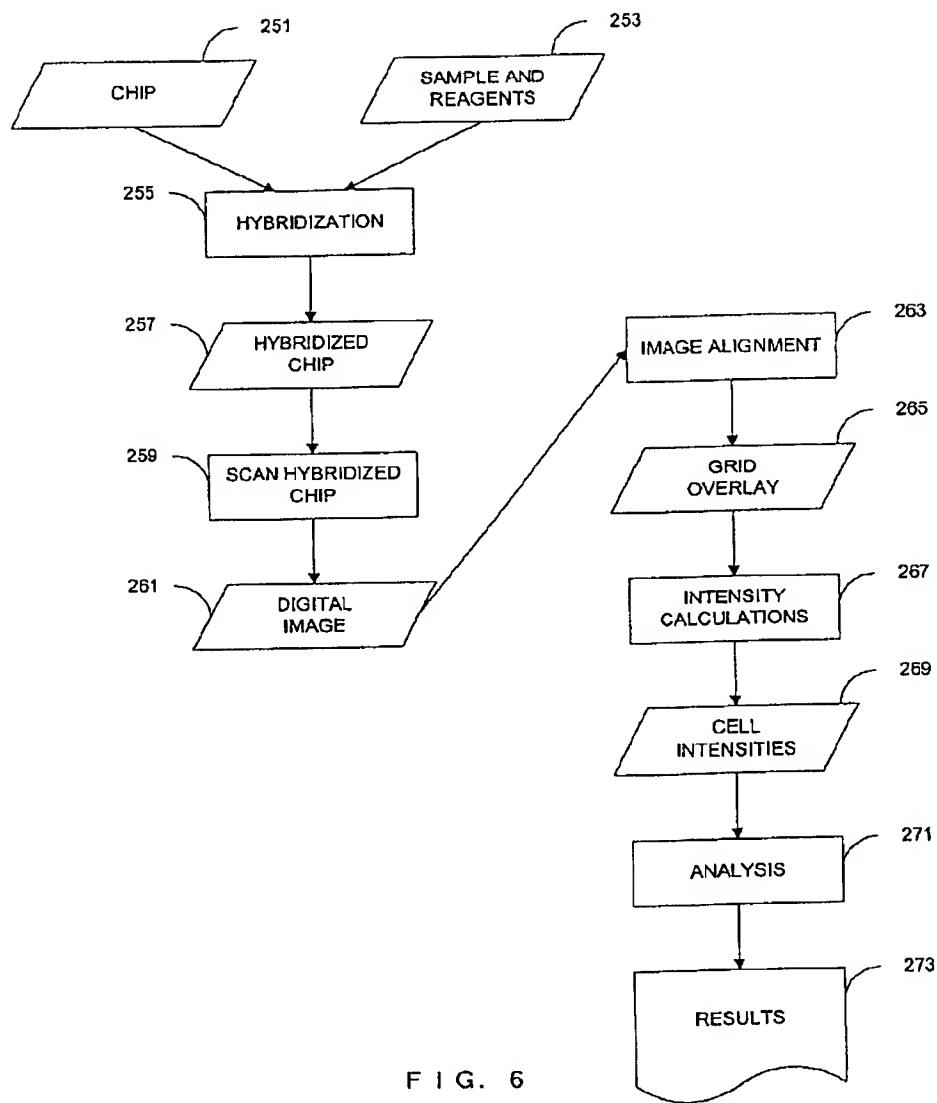


FIG. 6

(40) 2000-69998 (P2000-6(A))

整理番号= P A 5 7 C 6 8 3

ページ(6)

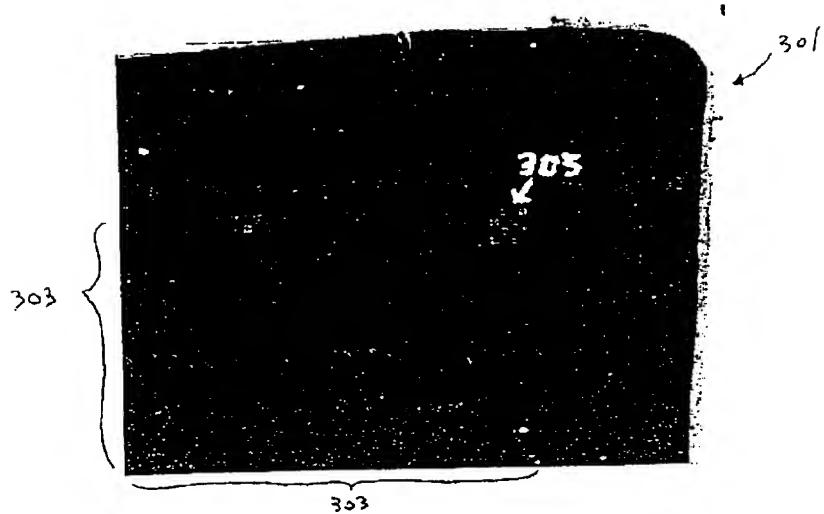


FIG. 7A

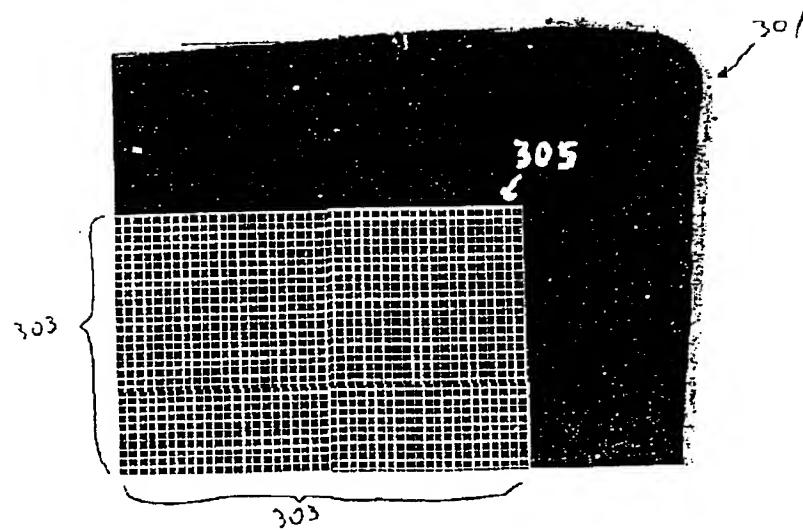
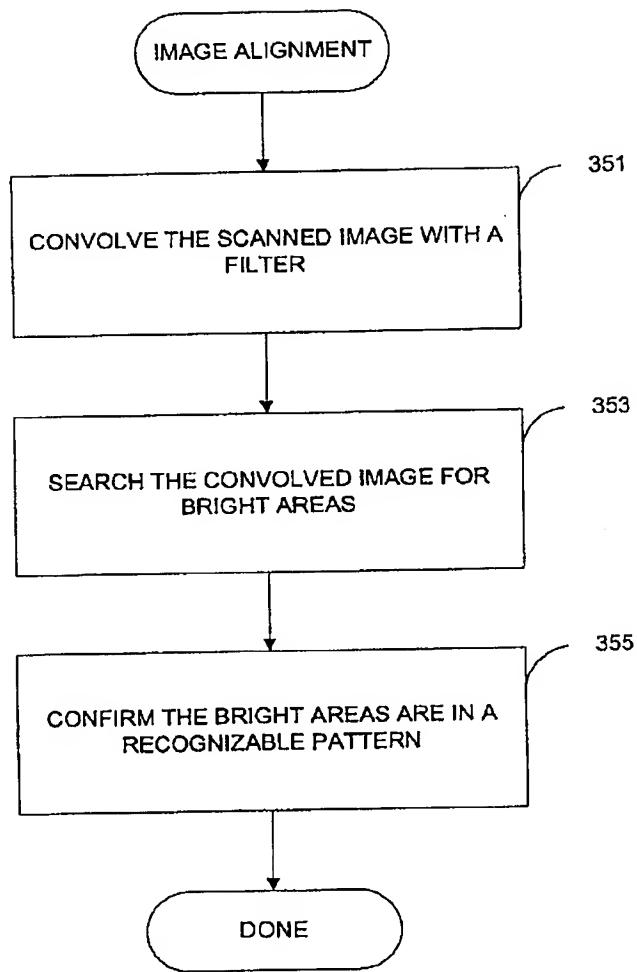


FIG. 7B

整理番号=PA57C683

ページ(7)



F I G. 8

(42)月2000-69998 (P2000-6寸級

整理番号= P A 5 7 C 6 8 3

ページ (8)

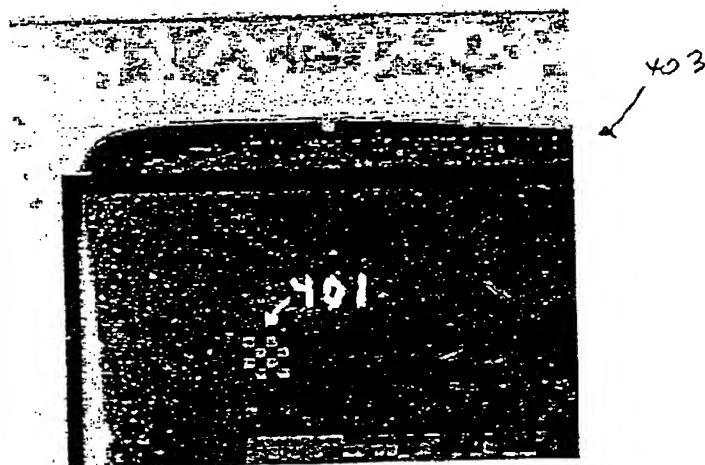


FIG. 9A

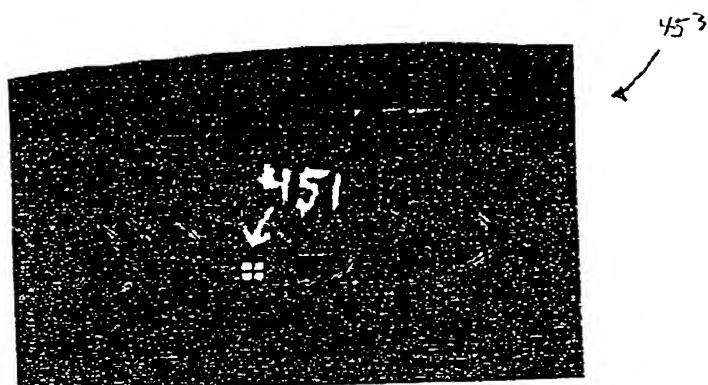
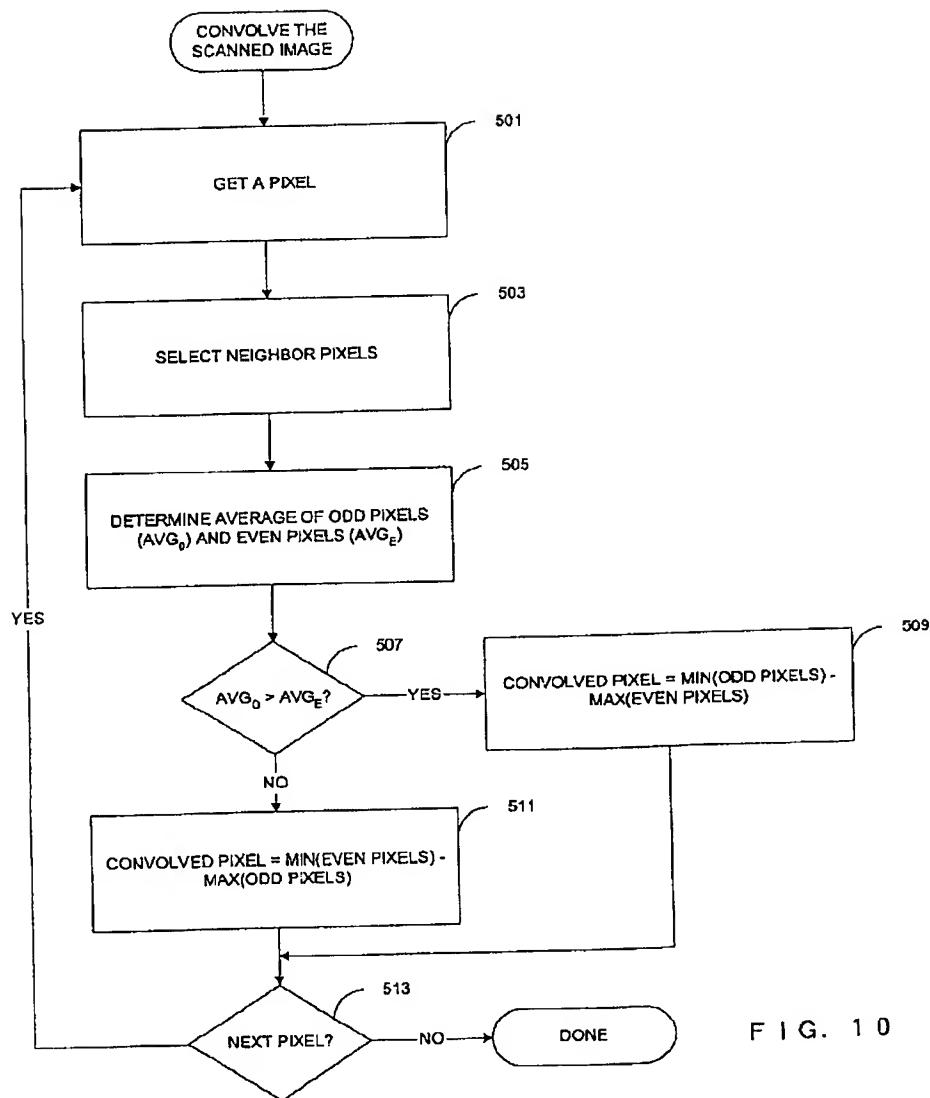


FIG. 9B



F I G. 10

(44) 2000-69998 (P2000-6A)

整理番号=PA57C683

ページ(10)

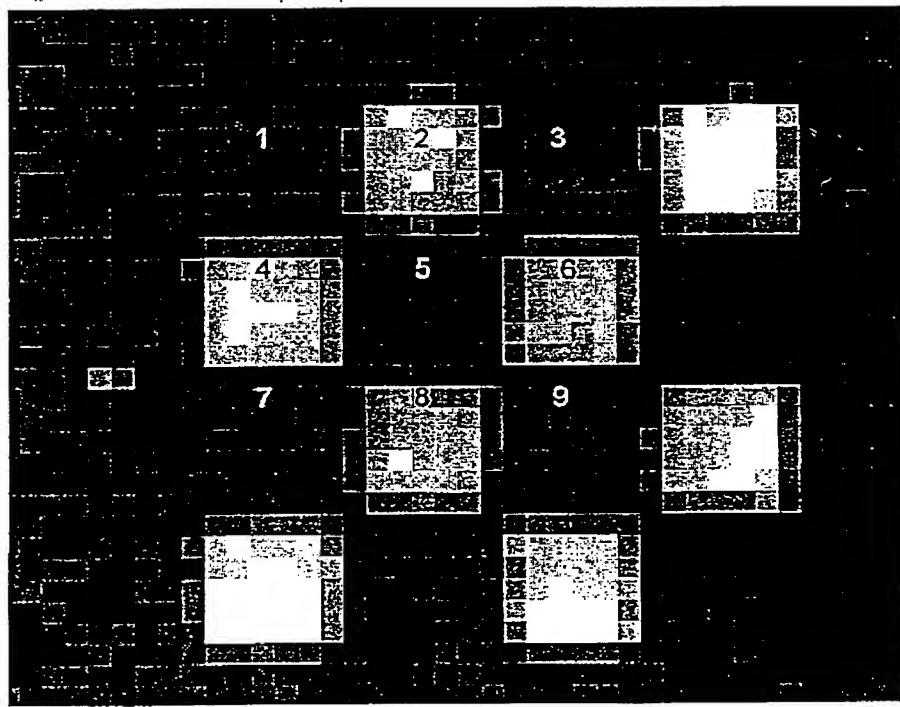


FIG. 11

(45) 2000-69998 (P2000-6#kA)

整理番号=PA57C683

ページ(11)

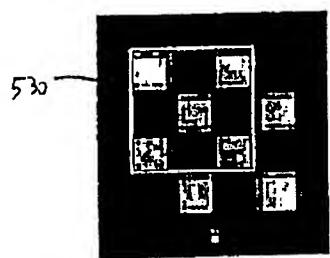


FIG. 12 A

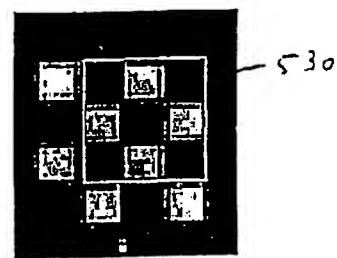


FIG. 12 B

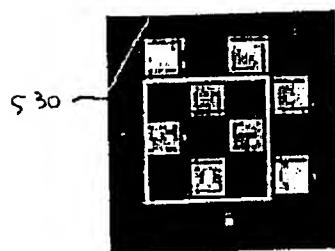


FIG. 12 C

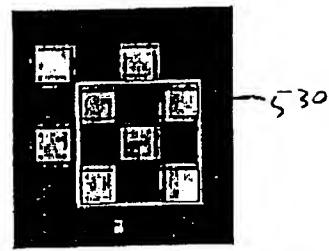


FIG. 12 D

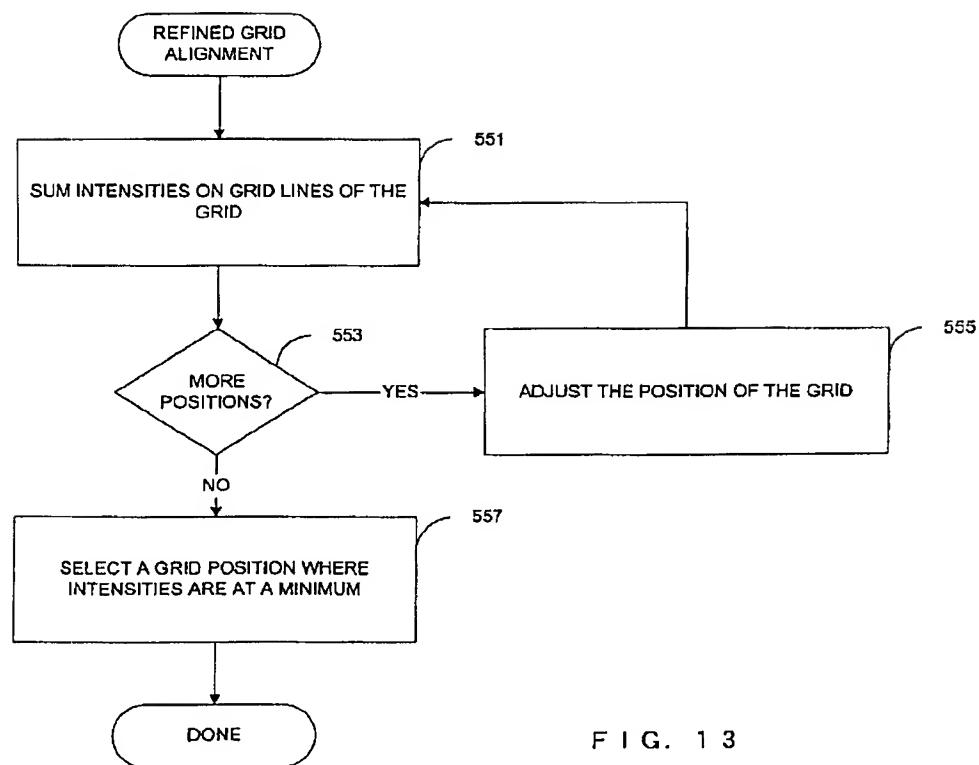


FIG. 13

(47) 月2000-69998 (P2000-60nA)

整理番号=P A 5 7 C 6 8 3

ページ(13)

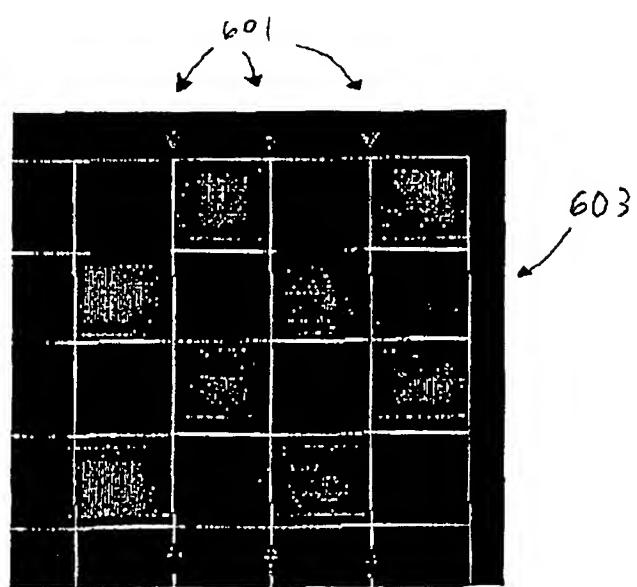


FIG. 14

1 Abstract

Systems and methods for aligning scanned images are provided. A pattern is included in the scanned image so that when the image is convolved with a filter, a recognizable pattern is generated in the convolved image. The scanned image may then be aligned according to the position of the recognizable pattern in the convolved image. The filter may also act to remove the portions of the scanned image that do not correspond to the pattern in the scanned image.

2 Representative Drawing

Fig.8